Application of Machine Learning to Improve the Interpretability of the Relationships between Silage Quality and Dairy Production

Haechan Mark Bong

Department of Animal Science
McGill University
Montréal, Québec, Canada

March 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science (Thesis)

Acknowledgements

First, I would like to thank my parents, Mekoung Lee and Wonsen Bong for giving me the opportunity to study abroad. Secondly, I thank my sister, Woori Zoe Bong for her continuous support throughout my education. I would like to also thank my close friends Zongpu Zed Zheng, Wei Peng Cai, David Santiago Pedreros Fonseca for constantly encouraging me to complete my degree.

During my research in McGill, my research was guided by my supervisor Kevin Wade. I would like to thank Prof. Kevin Wade for helping me to understand the fundamentals of machine learning and how it is applied to dairy industries in Québec. I would like to also thank my advisory committees: Dr. Roger I. Cue who taught me various data analysis approaches and how they could be applied to silage data; and Dr. Maxime Leduc and Dr. René Lacroix - experts in the field of data science - who shared their knowledge on the application of data science to silage and dairy science, which helped me significantly during my research.

Table of Contents

List of Tables

List of Figures

List of Appendices

ABSTRACT

While the impact of forage quality on dairy-milk production is clear, much work to date on such relationships as digestibility, fermentation characteristics and nutritional constituents with milk yield and composition have not easily been transferred to producers, given the complexity of analytical tests as well as their interpretation. With the growing interest in the field of precision dairy farming, the objective of this research was to provide better interpretability of the relationship between silage quality and milk production by performing an in-depth machine learning data analysis of forage variables such as digestibility, fermentation characteristics, and nutritional constituents.

Dairy production and forage data – characterized as either grass and legume silage or corn silage – were provided by Lactanet (Québec Dairy Herd Improvement). Since production data do not typically contain complete forage information, missing impact variables (like digestibility and fermentation characteristics) were predicted using supervised machine-learning algorithms; multi-input and multi-output (MIMO) regression was performed using a Meta-estimator with the Extra Tree algorithm as a base regressor for grass and legume silage, and a regressor chain based AdaBoost algorithm with an Extra Tree algorithm as the base regressor was used for corn silage. Factorial analysis was used to extract key silage quality characteristics (e.g., neutral detergent fiber digestibility, heat damage, rumen protein degradability, legume proportion, homolactic fermentation, length of initial fermentation, soil contamination, bad fermentation pattern), and a linear mixed effects model was used to estimate the effects of silage quality on average herd milk production and composition (fat, protein, milk-urea nitrogen and somatic cell score).

Increases in fermentation length, as well as the proportion of grass and legume mixture and corn silage or concentrate from feed, were associated with higher milk production, while the proportion of corn silage from feed showed an increasing trend in somatic cell count and milk urea nitrogen. Results from the research support the importance of forage quality on milk production and provide a method for completing missing forage values relative to time of feeding. Through this demonstration of the impact of forage quality, it is hoped to encourage better attention to this important input resource. The possibility of developing a decision-support tool for silage quality evaluation seems reasonable, thereby

helping to provide consultants and producers with a practical guide to improved dairy profitability from forage.

RÉSUMÉ

Alors que l'impact de la qualité du fourrage sur la production laitière est clair, de nombreux travaux à ce jour sur des relations telles que la digestibilité, les caractéristiques de fermentation et les constituants nutritionnels avec le rendement et la composition du lait n'ont pas été facilement transférés aux producteurs, compte tenu de la complexité des tests analytiques ainsi comme leur interprétation. Avec l'intérêt croissant pour le domaine de l'élevage laitier de précision, l'objectif de cette recherche était de fournir une meilleure interprétabilité de la relation entre la qualité de l'ensilage et la production de lait en effectuant une analyse approfondie des données d'apprentissage automatique des variables d'impact du fourrage telles que la digestibilité, la fermentation caractéristiques et constituants nutritionnels.

Les données sur la production laitière et le fourrage – caractérisées comme ensilage de graminées et de légumineuses ou ensilage de maïs – ont été fournies par Lactanet (amélioration du troupeau laitier du Québec). Étant donné que les données de production ne contiennent généralement pas d'informations complètes sur le fourrage, les variables d'impact manquantes (comme la digestibilité et les caractéristiques de fermentation) ont été prédites à l'aide d'algorithmes d'apprentissage automatique supervisés; une régression multi-entrées et multi-sorties (MIMO) a été réalisée à l'aide d'un méta-estimateur avec l'algorithme Extra Tree comme régresseur de base pour l'ensilage d'herbe et de légumineuses, et un algorithme AdaBoost basé sur une chaîne de régresseurs avec un algorithme Extra Tree comme régresseur de base a été utilisé pour l'ensilage de maïs. L'analyse factorielle a été utilisée pour extraire les principales caractéristiques de qualité de l'ensilage (p. ex. digestibilité des fibres au détergent neutre, dommages causés par la chaleur, dégradabilité des protéines du rumen, proportion de légumineuses, fermentation homolactique, durée de la fermentation initiale, contamination du sol, mauvais modèle de fermentation) et un modèle linéaire à effets mixtes a été utilisé pour estimer les effets de la qualité de l'ensilage sur la production et la composition moyennes du lait du troupeau (matières grasses, protéines, azote uréique du lait et score des cellules somatiques).

Augmentation de la durée de fermentation, ainsi que de la proportion de mélange d’ensilage d'herbes et de légumineuses et ensilage de maïs ou de concentré provenant des aliments pour animaux, étaient associées à une production laitière plus élevée, tandis que la proportion d'ensilage de maïs provenant des aliments pour animaux montrait une tendance à la hausse du nombre de cellules somatiques et de l'azote uréique du lait. Les résultats de la recherche confirment l'importance de la qualité du fourrage sur la production de lait et fournissent une méthode pour compléter les valeurs fourragères manquantes par rapport au moment de l'alimentation. Grâce à cette démonstration de l'impact de la qualité du fourrage, on espère encourager une meilleure attention à cette importante ressource d'entrée. La possibilité de développer un outil d'aide à la décision pour l'évaluation de la qualité de l'ensilage semble raisonnable, contribuant ainsi à fournir aux consultants et aux producteurs un guide pratique pour améliorer la rentabilité laitière des fourrages.

1. General Introduction

With over 12,000 dairy farms across Canada (Dairy Farmers of Canada, 2019), the dairy industry is a key contributor to the Canadian economy, where it ranks 3rd within Canadian agricultural activities (Producteurs De Lait Du Québec, 2019). According to the 2019 annual report of the Producteurs De Lait Du Québec 2019, Québec's dairy industry is the major producer of dairy products with 3.331 billion liters of milk from 4,877 dairy. Québec's high milk production is the result of multiple factors. Pluviometry and soil have always given Québec advantages in the production of grasses and pastures (Producteurs De Lait Du Québec, 2019). Also, Québec's abundant supply of high-quality silage has always been a key asset in its milk production (Producteurs De Lait Du Québec, 2019). It is also important to increase milk-fat, milk protein, lactose and milk yield to maximize profit and grow dairy industry in Québec. Price paid for milk production varies, but approximate pricing is listed in **Table 1.1**.

Table 1.1: Price for Milk Components

Milk Component	Price ($) / kg
Milk-fat	$10.80/kg + premium of $0.0145/kg for SNF/milk-fat ration of less than 2.35
Milk-protein	$7.8/kg
Lactose	$1.5/kg

SNF: Solid-Not-Fat

High quality silages are an important factor in cattle welfare and milk production. More specifically, silage quality is the key factor of cattle nutrition and feed intake, which affects their health condition, which then translates to the quality of milk production. Nutritive value, fermentation characteristics, pH, ammonia, and moisture all affect silage quality, and many of these factors have been discussed because they are modified by the ensiling process which in turn the ability of a cow to produce quality milk (Mertens *et al.,* 2009). The main determinants of high-quality silage are the concentrations of various nutritional constituents. Poor silage quality is usually due to ensiling crops with less-than-ideal ensiling management practices. This includes situations where crops are ensiled with improper moisture, maturity, packing, sealing, and feedlot management. Aerobically unstable or clostridial silages are the most common results when dealing with silage feeding problems, and butyric acid and nitrogenous proteolysis are the main result of clostridial activity. There are certain factors that are suggested as defining silage quality as illustrated in **Table 1.2**.

Table 1.2: Ideal Ensiled Silage and High Moisture Corn (HMC) Sample Fermentation Values (Seglar, 2003)

Test	Corn Silage	Grass Silage	HMC
*** Nutritional**			
Moisture Bunker / pile Stave / Bags Oxygen Free	(See footnote) below 63 – 72 60 – 68 50 - 60	 67 – 72 63 – 68 50 – 60	 26 – 32 26 – 32 22 - 28
ADF, % DM	23 – 30	30	3
NDF, % DM	46 – 50	55	9
NE – L, (mcal/lb DM)	0.68 – 0.7	0.75	0.93
NE – G, (Mcal/lb DM)	0.4 – 0.47		0.7 – 0.73
Crude Protein, % DM	7.1 – 7.9	18	10
Bound Protein % ADIN % TN	Less than 10 – 12	Less than 10 – 12	** N/A
Ammonia Nitrogen, % TN	Less than 10	Less than 15	Less than 10
*** Fermentation**			
pH	Less than 4	Less than 4.2	Less than 4.2
Lactic Acid % DM	Greater than 3	Greater than 3	Greater than 1
Acetic Acid % DM	Less than 3	Less than 3	Less than 1
Propionic Acid % DM	Less than 1	Less than 1	Less than 0.1
Butyric Acid % DM	Less than 0.1	Less than 0.1	Less than 0.1
Alcohol % DM	Less than 0.5	0	Less than 0.5
Microbial (* cfu/gm)**			
Yeast	Less than 100,000	Less than 100,000	Less than 100,000
Mold	Less than 100,000	Less than 1,000,000	Less than 100,000
Bacillus	Less than 100,000	Less than 100,000	Less than 100,000
*** Mycotoxin (****ppm)**			
Vomitoxin	0	0	0
All others	0	0	0

CORN SILAGE FOOTNOTE: Moisture is dependent on if crop is processed or not.

Legend for Table	Bunker	Stave	Sealed	Bag
Processed	67 – 72 %	63 – 68 %	55 – 65 %	60 – 68 %
Not Processed	63 – 72 %	60 – 68 %	55 – 65 %	60 – 68 %

* All Values Expressed as Dry Matter Basis
** N/A = Not Applicable
*** cfu / gm = colony forming units per gram silage or grain
**** ppm = parts per million of silage or grain

Generally, low pH leads to good preservation of silage, but is not enough, in itself, to measure true silage quality. The total mineral content of feedstuffs is called ash, where it is the inorganic material content. High ash levels may be indicative of excessive soil contamination coming in with the crop during harvest due to muddy or windy conditions (Seglar, 2003). Silage acids are also key indicators of silage quality. They consist of lactic acid, volatile fatty acids (VFAs) and ethanol. Lactic acid is critical for fermentation, and it is effective at reducing pH. Ideal silages usually have about 3 times more lactic acid (1 ~ 3%) than VFAs (acetic, butyric and propionic) (Seglar, 2003). Therefore, high lactic acid concentration generally suggests good silage quality. Acetic acid is usually found at less than 3% in silages, and anything over 3% suggests inefficient heterofermentative fermentation (Seglar, 2003). Propionic acid is usually found at less than 1% in normal silages (Seglar, 2003). Butyric acid should be less than 0.1% (Seglar, 2003). Elevated level of butyric acid indicates silage deterioration from secondary fermentation, which, in the presence of unpalatable nitrogenous end products such as amines and amides, may lead to a significant reduction in dry matter intake (DMI) and energy level of the silage (Seglar, 2003). The daily variation in silage dry matter (DM) and nutrient concentration is essential and can contribute significantly to maximizing silage quality. If silages have greater than 55% of DM, fermentation has a minor impact on the quality of silage (Linn, 1988). However, lower pH is desired for wetter silages to reduce proteolytic activity, which is one of the causes for low silage quality (Linn, 1988). Decrease in pH can be achieved through fermentation with an anaerobic environment, adequate substrate, and enough lactic acid bacteria (Muck, 1988). Studies have shown that pH values around 4 could significantly reduce proteolytic activity (Muck, 1988). More specifically, pH values for corn silage should be around 4 or less and for legume silage, it should be around 4 or slightly greater (Seglar, 2003). Silages are also affected by seasonal changes. In the case of grass silages, pH, crude protein (CP) and DM are mainly affected. There could be DM losses during wilting (Smith, 1954). When material is dry, it could be problematic during the ensiling process when preventing the temperature of the mass in the silo from rising quickly and resulting in reduced feed product (Smith, 1954). There is also an increase in protein content during the wet season (e.g., Rainfall) compared to dry season since grass is leafier and more mature during wet season (Smith, 1954). This does not mean that wet material is ideal for ensiling process since it could lead to spoilage. Therefore, farmers must ensure that the material is not too dry nor too wet for optimal ensiling process. Optimal moisture content is presented in **Table 1.2**.

Maximizing feed intake is crucial for optimal milk production. Selection of appropriate silage for cattle depends on their nutritional needs and environmental requirements, optimal particle size and highly digestible silage as some of the ways to increase feed intake. Combs (2015) states that digested fiber produced about 25% of the energy for milk production.

Despite extensive research efforts over the past 40 years, no generally accepted forage intake model has been developed (Huhtanen *et al*., 2008). Limited success in this field is, at least partly, due to complicated interactions between the animal and feed characteristics, and difficulties in distinguishing and quantifying these factors (Huhtanen *et al*., 2008). As a result, the objective of the present study was to conduct a data-driven analysis that would provide a general understanding of the relationships between the factors that define silage quality and its impact on milk production.

2. Literature Review

2.1 What Affects Milk Yield and Composition?

2.1.1 Environment

First, comfort of cattle is crucial for milk production. Discomfort can lead to reduced intake and health concerns. There are several ways to optimize cattle comfort. To list a few, using a stocking rate (the number of specific kinds and classes of animals grazing or using a unit of land for a specific time period) at 80 to 85% of capacity, keeping cattle in a fresh cow group for 14 to 21 days can help bring comfort for cattle (Litherland, 2018). In addition, providing 30 to 36 inches of bunk space per cow, reducing social stress (especially for the first calf heifers) and preventing cows from separating from the normal herd mates can significantly reduce cattle stress (Litherland, 2018). Setting an ideal temperature for dry and lactating cows is also essential. Climate change has increased the overall temperature causing an imbalance between metabolic heat production inside the animal body and its dissipation to the surroundings results in heat stress (HS) under high air temperature and humid climates (Das *et al*., 2016). The foremost reaction of animals under thermal weather is an increased respiration rate, rectal temperature, and heart rate (Das *et al*., 2016). It directly affects feed intake thereby, reduces growth rate, milk yield, reproductive performance, and can cause death in extreme cases (Das *et al.*, 2016). Dairy breeds are typically more sensitive to HS than meat breeds, and higher producing animals are, furthermore, susceptible since they generate more metabolic heat. Heat stress suppresses the immune and endocrine system thereby enhances susceptibility of an animal to various diseases. Bouraoui *et al.* (2002) observed lower milk-fat and milk-protein in the summer season. Milk-fat in the summer tends to be lower in palmitic acid relative to stearic and octadecanoic acids than milk-fat from the same cows during the winter (Christie, 1979). Therefore, a cooling system is often necessary.

2.1.2 Health

Milk production is also directly related to the health of cattle since it results from reduced intake and diseases. Monitoring (Body Condition Score) BCS is important, where BCS is a simple technique to assess how thin or fat a cow is on a scale of 1 to 5 with increments of 0.25, where 1 is extremely thin and 5 is extremely fat (Das *et al*., 2016). According to Das *et al*. (2016), other important part of cattle health is the hoof health. At the same time, attention to subclinical milk fever is necessary. This is caused by low blood calcium, which is a result of multiple factors such as ketosis, high somatic cell count, delayed uterine involution, metritis, reduced feed intake and milk yield (Das *et al*., 2016). Mastitis (inflammation of the udder) is common for dairy cows and known to generally cause a decline in milk-fat percentage and a change in milk-fat composition (Kitchen, 1981). The general effect of mastitis is to impair milk synthesis and loosen the connections between cells, thereby increasing permeability of blood constituents (Jenness, 1985). As a result, milk-proteins synthesized in the mammary gland (caseins, beta-lactoglobulin, and alpha-lactalbumin) decrease (Kitchen, 1981), whereas blood serum proteins (whey proteins) increase (Kitchen, 1981). The decrease in fat percentage, however, is less (~10%) than that observed for lactose or casein (~15%) (Linn, 1988). The percentages of calcium and phosphorus in milk also decline with mastitis infections and lowers casein levels since both ions are complexed with casein micelles (Kitchen, 1981). Kitchen (1981) also added that mastitis increases the percentages of sodium and chloride in milk and decreases the percentage of potassium. Bacterial infection of the udder results in damage to the ductal and secretory epithelium and increases the permeability of blood capillaries (Linn, 1988). Thus, sodium and chloride, which are higher in blood, pour into the lumen of the alveoli, and to maintain osmolality, potassium is decreased proportionally (Linn, 1988).

Growth hormone affects synthesis of fatty acids in the mammary gland and uptake of pre-formed fatty acids from the blood, depending on dose level and energy balance of the cow (Linn, 1988). The hormone requirement for milk synthesis and secretion is prolactin, adrenocorticotrophic hormone, and estrogens, and the relative absence of progesterone (Linn, 1988). Administration of exogenous growth hormone has generally shown to increase in milk yield without significant changes in composition (Bauman *et al*., 1985). However, Bauman *et al*. (1985) observed a slight decrease in milk-protein percentage and an increase in alpha-lactalbumin as a percentage of total milk-protein with increasing dosage levels (0 to 100 IU/day) of bovine growth hormone. It is ideal to have feed additives such as rumen-protected choline to improve cattle health (Das *et al*., 2016).

2.1.3 Genetics

Breeds differ in total milk-protein percentage and type of milk-protein produced (Linn, 1988). Depending on the dairy breed, the milk yield and its composition varies significantly. According to the Canadian Dairy Information Centre's milk recording by breed (**Table 2.1,** Canadian Dairy Information Centre, 2015), Holstein is the predominant breed in Canada (93%).

Jersey and Guernsey cattle have the highest percentages of total protein, casein, and whey (Linn, 1988). Variability of the major protein fractions within breeds has also been reported (Rolleri *et al*., 1956), with Holstein milk containing less of the major caseins and more gamma-casein than milk from other breeds. Therefore, genetic selection would increase the percentage of protein in milk 0.075 percentage units, but decrease milk yield 231 pounds (Linn, 1988). Joint selection for milk yield, protein, and fat is recommended if the desired result is increased yield of protein and fat (Gaunt, 1980).

Table 2.1: Average Milk Production by Breed in Canada Canadian Dairy Information Centre (2015).

Breed	**Records**	**Milk (Kg)**	**Fat %**	**Protein %**
Ayrshire	8146	7.78	4.11	3.37
Brown Swiss	1731	8.40	4.20	3.49
Canadienne	194	5.75	4.34	3.57
Guernsey	370	6.76	4.69	3.43
Holstein	287223	10.10	3.87	3.19
Jersey	11334	6.61	5	3.80
Milking Shorthorn	347	6.81	3.94	3.28

2.1.4 Nutrition

Without proper nutrition, dairy cows are not able to provide optimal milk production. Diets for today's high-producing dairy cows are typically higher in energy from readily fermentable carbohydrates than fats and feeding of these diets often causes a condition known as low-milk-fat syndrome (Linn, 1988). Characteristics of low-milk-fat syndrome are a reduction in milk-fat percentage (up to 60%) and changes in milk-fat composition (increase in C_{18} polyunsaturated and monounsaturated acids and decrease C_{160} and C_{180} fatty acids) (Christie, 1979).

Shaver *et al*. (1986) has shown that milk-fat percentages are higher from cows fed a 60:40 silage to grain diet at 2.93% of body weight than at 3.75% of body weight. Declines in milk-fat percentage with high-grain feeding are accompanied by a change in milk-fatty acid composition from saturated fatty acids to more unsaturated acids, especially those containing

16 carbons or less (Banks *et al*., 1983). The type of silage and its effect on milk-fat percentage are influenced by silage particle size, maturity, and fiber content (Linn, 1988). Finely ground silages result in higher levels of propionate being produced during rumen fermentation than silages of adequate particle size (Sutton, 1980). Woodford *et al*. (1986) has shown that a mean silage particle length of 0.64 cm or more is needed to keep rumen molar percentage of propionate below 25 and milk-fat above 3.6%. Stage of silage maturity is also an important factor in the supply of adequate fiber in the diet. More immature alfalfa hay was required in the diet to obtain maximum production of 4% fat-corrected milk than when mid- or late-bloom alfalfa hay was fed (Kawas *et al*., 1983). Sutton (1985) reported that the lower ruminal degradability of corn compared with that of barley would result in the production of milk with a higher fat percentage. DePeters and Taylor (1985) confirmed that barley-based concentrates tend to depress fiber digestibility, resulting in lower ruminal acetate to propionate ratios and lower milk-fat percentages than those with corn-based concentrates. The higher digestion of barley in the rumen produces more propionate and results in less starch being presented to the lower digestive tract for conversion to glucose than with corn (Linn, 1988). However, the increased production of propionate in the rumen from barley appeared to stimulate milk yield more than glucose derived directly from corn in the lower digestive tract (Linn, 1988). Sutton (1980) suggests that processing of grains such as grinding, rolling, heating, steam flaking, and pelleting increases digestion of the starch in the rumen and produces effects like those reported above for barley.

Increasing butyric acid production in the rumen should also help to maintain or increase milk-fat percentages, and Sutton (1980) suggested that beet pulp is a promoter of butyric acid production in the rumen. Other carbohydrates such as whey (Casper and Schingoethe, 1986), sucrose, and lactose (Sutton, 1980) have been evaluated as sources of soluble carbohydrate to prevent milk-fat depression. Linn (1988) added that intraruminal infusions of acetic acid consistently increase milk yield, lactose yield, and milk-fat yield, whereas infusions of propionate reduce milk-fat yield. In addition, glucose infusions, either intraabdominal or intravenous, increase milk yield and decrease milk constituent percentages (Linn, 1988).

Additives such as buffers and methionine hydroxy analog have been used to promote increases in milk-fat percentage (Linn, 1988). Lundquist *et al*. (1983) also states that cows in early lactation, fed with high-concentrate diets, were shown to benefit from the inclusion of the methionine hydroxy analog in their rations. Linn (1988) added that feeding of 25 grams of methionine hydroxy analog daily during the first 120 days of lactation increased milk-fat 0.35 percentage units. Buffers are compounds used to raise rumen pH through the neutralization of volatile fatty acids. Protected polyunsaturated fatty acids appear to be the most promising for consistently increasing milk-fat percentage and altering milk-fat composition (Linn, 1988). However, other modes of action have been indicated for the group of compounds commonly alluded to as buffers (sodium bicarbonate, potassium bicarbonate, limestone, magnesium oxide, and bentonite) (Chalupa and Schneider, 1985). Chalupa and Schneider (1985) added that in general, bicarbonates have been effective in maintaining or increasing milk-fat percentages of cows fed high-grain diets, especially when corn silage was the main silage source. Magnesium oxide has also been shown to help prevent milk-fat percentage depression (Linn, 1988). Intake of digestible organic matter (OM) increased ($P < 0.001$) with crude protein (CP) concentrate supplementation, but the response tended to diminish at high levels of supplementation (Nousiainen and Rinne, 2009).

Linn (1988) mentioned that protected oilseeds or oils rich in linoleic acid (sunflower, corn, and soybean) produce rapid increases in the linoleic acid content of milk-fat when fed. The increases in linoleic acid content are generally associated with declines in myristic, palmitic, and oleic acids (Linn, 1988)**.** Linn (1988) added that feeding of protected saturated fats – the most common source being tallow – generally invokes the same response in increase of milk-fat percentage as feeding of protected polyunsaturated fats. However, protected hydrogenated soybean oil has decreased the milk-fat percentage (Banks *et al*., 1983).

2.2 Impact of Silage Quality

2.2.1 Impact of Silage Quality on Feed Intake

Several studies have shown increased intake with corn silages compared to legume silages and legume silages compared to grass silages. (O'Mara *et al*., 1998). Some reasons that one type of silage increases intake compared to other are due to its fiber content and digestibility, particle size and variability (Mertens *et al*., 2009). Fiber is a nutritional term that is defined as either the indigestible or slowly digesting fraction of feeds. Digested fiber produces about 25% of the energy for milk production (Combs, 2015). A simple summative equation demonstrates that fiber content and its digestibility are the major factors that affect the total DM digestibility of feeds (Mertens *et al.,* 2009). Fiber digestibility is mainly dependent on lignin content due to its indigestibility in the rumen. Neutral detergent fiber (NDF) can separate feeds into almost completely digestible neutral detergent soluble fiber and NDF that varies in digestibility. In general, grasses have a higher proportion of NDF compared with legumes, but the proportion of lignin in total NDF is higher in legumes (Beever *et al*., 2000). Intake is often the limiting Factor in dairy cow productivity and fiber can often limit the intake of silages. Particle size of silage also has an impact on feed intake. Optimizing silage particle size is important because excessively long particles increase the necessary chewing to swallow a bolus of feed, thereby increasing eating time (Grant *et al*., 2018). Under competitive feeding situations, excessively coarse or lower fiber digestibility silages may limit dry matter intake (DMI) of lactating dairy cows due to eating time requirements that exceed available time at the feed bunk (Grant *et al*., 2018). With grass silages, Van Soest (1994) observed the presence of higher proportion of indigestible NDF (iNDF) in regrowth silages, probably reflects warmer growing conditions. This has an impact of lower intake of regrowth silages compared to primary growth silages (Huhtanen, Rinne and Nousiainen, 2007). Pang (2019) added that the lower intake potential of regrowth silage could also be due to other factors such as silage microbiological quality, increased amount of decomposing and infected leaf material, which possibly contributed to the taste, smell or palatability of the silage.

There are several silage components that are known to reduce feed intake: for example, condensed tannins (Provenza *et al*., 1990) and glucosinolates (Duncan and Milne, 1993). Impaired silages with high VFAs can result in less nutritious and, therefore, less desirable feed for animals. In the case of corn silages, production of butyric acid during clostridial activity is a primary concern. More specifically, butyric acid is known for reducing feed intake in ruminants (Muck, 1988). Mycotoxin-contaminated diets are known to reduce feed intake (Zain, 2011). High concentration of acetic acid is also know to reduce intake due to increased osmotic pressure of ruminal contents (Forbes *et al*., 1992). Buchanan-Smith (1990) added that reduced dry matter intake (DMI) due to acetate can be affected by palatability. High concentrations of ammonia are also known to reduce feed intake (Huhtanen *et al.*, 2007). Auerback *et al.* (1998) reported a lack of appetite in cattle herds fed silages containing 0.2 to 1.5 mg of roquefortine C/kg in northern Germany.

Minson (1990) suggested that, for feeds with a CP content of less than 62g of CP per kg DM, fiber digestion is inhibited, and he reports a number of trials in which intake of silages increased by 14 - 77% following provision of supplementary protein. It is also observed that changes in CP composition through protein degradation can reduce feed intake (Südekum and Eisner, 2009). Where ammonia N concentration limits microbial fermentation, supply of N to the microorganisms increases organic matter (OM) digestion in the rumen, which increases breakdown and rate of passage of a poor-quality silage, thereby removing the physical constraint and allowing the animal to consume more feed (Romney *et al*., 2000).

2.2.2 Impact of Silage Quality on Milk Production

Fermentation of silages has a direct influence in determining silage quality. When consumed by dairy cows, nutrients profile from the silage affects milk yield and composition. According to studies, corn and legume silages produce higher milk production and milk-protein concentration compared to grass silage (O'Mara *et al*., 1998). This is potentially due to their higher digestibility. Milk-fat content decreased when there was an increase in lactic acid and volatile fatty acids (VFA). Increases in lactic acid concentration reduced the milk-fat content more at high than low concentrations compared to VFA which caused greater changes at low than high concentration (Huhtanen *et al*., 2003). Like milk-fat content, milk-protein content also decreased as lactic acid and VFA increases. Increases in individual or VFA had more impact on depressing milk-protein content than lactic acid (Huhtanen *et al*., 2003). Effects of silage fermentation characteristics on milk-fat yield has similar patterns as milk-fat content. Milk-fat yield decreased when lactic acid and VFA increases. Like milk-protein content, milk-protein yield decreased as lactic and VFA increases. Silage pH did not have an impact on milk-protein yield (Huhtanen *et al*., 2003). Using multiple regression, it was observed that VFA and lactic acid explained the variation of energy corrected milk (ECM), milk-fat content, milk-fat yield, milk-protein content and milk-protein yield better than other components of silage

(Huhtanen *et al*., 2003). The models suggest in general, that propionic and butyric acid decreased milk yield more than acetic acid.

Milk-fat percentage is related positively to rumen molar percentages of acetic and butyric acids and negatively to that of propionic acid (Linn, 1988). Davis (1978) reported that rumen molar percentage of propionate must be above 25 before a highly significant negative relationship between milk-fat percentage and propionate exists. Sutton (1980) estimated that 60 percent of the variations observed in milk-fat percentage could be accounted for by changes in the molar proportion of propionate in the rumen. A positive relationship exists between the molar ratio of acetate to propionate and milk-fat percentage. A linear increase in milk-fat percentage occurs as the ratio of acetate to propionate increases up to 2.2 (Davis, 1978). Above a ratio of 2.2 there is little change in milk-fat percentage. Thus, diets that increase propionate production have the greatest effect on milk-fat percentage when considering total acids.

The daily amount of NDF needed was estimated to be 1.2% of body weight. Mertens (1985) recommended a minimum of 28 percent NDF and about 18% acid detergent fiber (ADF) in diets to maximize milk production and fat percentage. The average increase in milk and energy-corrected milk yield was 0.30 and 0.37 kg per 10-unit increase in silage organic matter digestibility (*D*-value), respectively (Pang *et al*., 2019). Milk-protein concentration increased, and fat concentration tended to increase with enhanced silage *D*-value (Pang *et al*., 2019). In the evaluation of Pang *et al*. (2019) evaluation, every 10-unit increase in silage *D*-value increased milk-protein yield by 12.9g/day, which was close to the 11.0g/day reported by Rinne (2000). The increase in milk-protein yield per unit increment in DMI was 52g/kg, which agreed with the 50g/kg in Kuoppala *et al.* (2008). These combined factors result in lower milk production in cows fed regrowth silages compared with cows fed primary growth silages (Kuoppala *et al.,* 2008).

High ethanol content has been occasionally observed in high-dry-matter grass silages (Kalac, 2011). In such cases, lactic acid fermentation is limited, and ethanol is the main product of fermentation (Kalac, 2011). The milk acetone concentration was doubled when ethanol was fed (Kalac, 2011). High levels of ethanol in silages can decrease milk yield but increase milk-fat and protein concentrations and induce milk off-flavor (Randby *et al*., 1999).

The carotenoids contribute in the oxidative stability of milk (Kalac, 2011). Despite a higher antioxidative capacity of milks from cows fed grass silage, lipid oxidation was higher as compared to milks from cows fed corn silage (Kalac, 2011). To the contrary, the milk from cows fed corn silage was more vulnerable to protein oxidation (Havemose *et al*., 2004). Corn silages have been a poorer source of carotenoids than silages of other crops, especially if prepared from corn damaged by frost (Kalac, 2011).

Ensiled grasses and legume silages seem to contain higher levels of available tocopherols than corn silage (Kalac, 2011). The grass-red clover silage showed to be a richer source of available tocopherols than corn silage (Kalac, 2011). The concentrations of α-tocopherol were 0.85 and 0.38 mg l^{-1} and those of γ-tocopherol 0.03 and 0.01 mg l^{-1} in the milk of dairy cows fed grass-clover silage or corn silage, respectively (Havemose *et al*., 2004). More rapid losses of α-tocopherol and formation of oxidative products were observed in milk from dairy cows fed diets based on red clover or lucerne silages than from those fed grass silages (Kalac, 2011). The increased oxidative deterioration of milk from cows fed red clover silage was avoided by vitamin E supplementation (Al-Mabruk *et al*., 2004).

Contamination of raw cow milk by *Listeria monocytogenes* has been linked to the occurrence of high levels of *Listeria monocytogenes* in silage (Sanaa *et al*., 1993). *Listeria monocytogenes* occurs at low numbers in raw cow milk (Driehuis *et al.*, 2018). Surveys of the prevalence of Listeria *monocytogenes* in bulk tank milk from dairy farms in the United States, New Zealand, France, and Belgium showed that 2.9 to 6.3% of the samples were positive (tested for presence in samples of 25 g (Desmasures *et al*., 1997; de Reu *et al.*, 2004; Van Kessel *et al.,* 2011; Marshall *et al*., 2016). Driehuis *et al*. (2018) reported that *Listeria monocytogenes* is sensitive to heat inactivation and is effectively inactivated by the pasteurization of milk.

Feeding is an effective way to modify the sensory quality of dairy products, even in the case of milk bulk tanks mixtures (Kalac, 2011). Also, soil is a major source of B. cereus spores in silage. B. cereus can cause atypical appearance and small which then spores dairy products. Dairy plants could market different milks, which would be of specific composition (Agabriel *et al*., 2007).

2.2.3 Impact of Silage Quality on Cattle Health

Good health of dairy cows is crucial for optimal feed intake and milk production. There are multiple factors of silage characteristics that contribute to cattle health. To list some recommendations for cattle health improvement, it is suggested to maintain DMI of 12.7kg to 14.5kg per day and avoid overfeeding energy (Das *et al*., 2016). There are over 400 mycotoxins and ZEA is one of the main mycotoxins formed in silage (Kalac, 2011). ZEA – a macrocyclic β-resorcyclic acid lactone – is an estrogenic metabolite produced by several species of Fusarium such as *F. graminearum, F. roseum, F. culmorum, and F. crookwellense* (Kuiper-Goodman *et al*., 1987; Saeger *et al*., 2003). It also leads to some serious health concerns such as infertility, and hyperestrogenism in cattle (Ogunade *et al.*, 2018). A survey from Rodrigues and Naehrer, 2012 reported that approximately 45% of 7,049 livestock feed samples collected from the Americas, Europe, and Asia contained ZEA, with an average concentration of 233 µg/kg. Its content can be reduced by the activity of both some lactic acid bacteria in silage and the rumen microflora (Kalac, 2011). The degradation products resulting from proteolysis may impair animal health (Hoedtke *et al*., 2010). Determining amine concentrations in silage may help to indicate undesirable changes in silages and could prevent possible toxicity for livestock (Křížek, 1991). Křížek 1991 discussed that many liver and kidney disorders related to the detoxification and catabolism of biogenic amines (Scherer *et al.*, 2015). Microbial hazard such as clostridium botulinum concentrations in silage have been associated with botulism in cattle. A high initial concentration of clostridium botulinum spores in silage in combination with poor silage fermentation conditions can promote the growth of clostridium botulinum in silage. Generally, exposure to low numbers of *C. botulinum* or spores of this microorganism is not harmful (Driehuis *et al*., 2018). However, any factors that induce multiplication of C. botulinum must be avoided because of the extreme toxicity of the botulinum toxin (Driehuis *et al*., 2018). With the elevation of pH level, other microbial hazards such as *L. monocytogenes*, Shiga toxin-producing *E. coli*, and molds in silage may also encourage survival and growth of *M. bovis*, the bacterium that causes bovine tuberculosis (Driehuis *et al*., 2018). Listeriosis is often considered a food-borne disease of ruminants, with silage being the main feed source (Driehuis *et al*., 2018). A causal relationship has been shown between feeding poor-quality silage and the prevalence of listeriosis in cattle, sheep, and goats (Fenlon, 1988; Wiedmann *et al*., 1997; Ho *et al*., 2007). *L. monocytogenes* from silage survives passage through the animal's gastrointestinal tract and is shed in the feces (Driehuis *et al*., 2018). This has not only been observed for cattle, sheep, and goats, with clinical signs of listeriosis, but also for asymptomatic carriers of *L. monocytogenes* on farms with and without outbreaks (Unnerstad *et al*., 2000; Nightingale *et al*., 2004; Vilar *et al*., 2007). In these animals, L. monocytogenes primarily causes encephalitis and uterine infections, the latter causing late-term abortions (Driehuis *et al*., 2018). In addition, *L. monocytogenes* can cause eye infections in ruminants (silage eye) because of direct contact with silage (Erdogan, 2010). Important features of *L. monocytogenes* are that it can grow over a wide range of temperatures (0 – 45°C), salt concentrations (up to 12%), and pH (4.3 – 9.6) (Van der Veen *et al*., 2008; Gandhi and Chikindas, 2007). Due to its high tolerance to stressful conditions, *L. monocytogenes* is capable of survival for extended periods in environments in which it is unable to grow – e.g., well-preserved silage (Driehuis *et al*., 2018). Survival of *L. monocytogenes* in silage is determined to a great extent by the degree of anaerobiosis and the pH (Driehuis *et al*., 2018). Donald *et al*. 1995 showed that the population of *L. monocytogenes* rapidly declined under strictly anaerobic conditions when it was added to grass at ensiling, whereas oxygen tensions of 0.5% (vol/vol) and higher prolonged survival. Therefore, poor compaction before ensiling or air ingress during the fermentation can promote the growth or persistence of this pathogen. Growth of *L. monocytogenes* in silage is also associated with aerobic deterioration problems (Driehuis *et al*., 2018). The combination of the presence of oxygen and relatively high pH in aerobically deteriorated silage favors growth of *L. monocytogenes*. Silages with a greater likelihood of aerobic surface spoilage (e.g., silage with a low packing density or inadequate sealing and baled silages), are most susceptible to contamination by *L. monocytogenes* (Fenlon *et al*., 1989). Studies by Ryser *et al*. (1997) and Vilar *et al*. (2007) have shown that the incidence of *L. monocytogenes* in silage increases with increasing pH. *L. monocytogenes* was detected in concentrations in excess of 10^6 cfu/g in moldy surface layers of big bale grass silages (Fenlon, 1986).

Mycotoxins from silage can also affect milk production. Fumonisin is another mycotoxin that has been identified to be harmful to dairy cows. Mathur *et al*. (2001) states that fumonisin could be nephrotoxic to calves when fed 1,000 µg/kg BW of the toxin. Similar results were observed in beef calves supplemented with 148 mg/kg of total fumonisin in the diet for 31 d (Osweiler *et al*., 1993). Zearalenon (ZEA) is also reported to decrease milk production Coppock *et al*., 1990). Symptoms such as reproductive disorders and mastitis were detected in cattle herds fed silages containing 0.2 to 1.5 mg of roquefortine C/kg in northern Germany (Auerbach *et al*., 1998). Also, paralytic effects were reported in cows fed 4 to 8 mg of roquefortine C/kg (Haggblom, 1990).

Hazards from plant toxins include pyrrolizidine, tropane and tropolone alkaloids, phytoestrogens, prussic acid, and mimosine compounds that exist naturally in certain plant species that may contaminate silages at harvesting (Driehuis *et al*., 2018). Another group of toxins belonging to this category are ergot alkaloids, which are produced by endophytic fungal species in silages such as tall fescue grass (*Festuca arundinacea*), sorghum, and ryegrass (*genus Lolium*). Chemical and microbiological hazards are associated with poorly fermented silages, which can be avoided by using proper silage-making practices and creating conditions that promote a rapid and sufficient reduction of the silage pH and prevent aerobic deterioration (Driehuis *et al*., 2018). *Aspergillus fumigatus* is considered a health hazard not only because of the mycotoxins that are potentially produced by this mold in silage, but also because inhalation of spores of this mold can cause disease (aspergillosis) in humans and animals (Driehuis *et al*., 2018). Feeding silage or hay that is contaminated with *A. fumigatus* to cattle can cause bovine aspergillosis (Smith and Lynch, 1973; Sarfati *et al*., 1996). In immune-compromised individuals, this mold can cause severe infections (invasive aspergillosis; Dagenais and Keller, 2009). Puntenney *et al*. (2002) suggested that *A. fumigatus* is a risk Factor for hemorrhagic bowel syndrome in dairy cows; Kallela *et al*. (1984) observed serious fertility problems in heifers fed ensiled red clover containing high levels of estrogenic isoflavone. In 1994, outbreaks of ergot toxicosis in Africa were reported in cattle that ingested diets contaminated with ergotized annual ryegrass seed; approximately 2,646 dairy cows had reduced milk production, weight loss, and reduced fertility (Schneider *et al*., 1996). Two dairy herds in South Africa, that consumed corn silage, developed hyperthermia and experienced 30% loss of milk production; a novel endophyte, *Claviceps cyperi*, and up to 0.98 mg/kg of total ergot alkaloids were discovered in the corn silage (Naudè *et al*., 2005). In addition, Lean 2001 reported that lactating dairy cows consuming perennial ryegrass silage, containing 1.78 mg/kg of total ergovaline, experienced reduced reproductive performance, increased incidence of mastitis, and decreased milk yield. Blaney *et al*.

(2000) summarized several cases and suggested that signs of ergot alkaloid toxicity include feed refusal and severe declines in milk yield.

Production of nitric oxide and nitrogen dioxide gases in silage is also a serious risk to livestock and human health and the respiratory hazard, known as silo-fillers' disease has been recognized for many years (Grayson, 1956). The reduction of NO_3^- to nitric oxide, a colorless gas, is followed by its oxidation by exposure to air, which is then turned into nitrogen dioxide, a yellow to reddish-brown gas with an irritating odor that is heavier than air and stays close to the ground level or to the sides of the silo (Driehuis *et al*., 2018). Nitric oxide and nitrogen dioxide react with water in air to form nitrous and nitric acid gases, respectively (Driehuis *et al*., 2018). Inhalation of these gases damages lung tissue and causes respiratory distress leading to asphyxiation (Driehuis *et al*., 2018). O'Kiely *et al*. (1999) described an incident in which 10 calves and the farmer suffered severe respiratory distress 24h after an adjacent silo had been filled with grass. A brown haze of nitrogen dioxide gas was seen in the building next to the silo. O'Kiely *et al*. (1999) added that samples taken from the front of the silo beneath the polyethylene film covering revealed an intensely yellow colored layer beneath normal, green-colored silage. Features of the yellow silage were very high CP, extremely low pH, low concentrations of fermentation acids, low buffering capacity, and reduced digestibility in vitro, indicative of nitric acid contamination (O'Kiely *et al*., 1999). The growth of clostridia in silage can be also associated with significant livestock health issues as a result of absorption of butyric acid from rumen into blood can increase the risk of clinical ketosis (acetonaemia) (Oetzel, 2007).

2.2.4 Existing Literatures on the Relationship Between Silage Quality and Milk Production

Although various studies were conducted over the past 40 years and there is still an absence of a generally accepted feed intake model for dairy cows which optimizes milk production (Huhtanen *et al*., 2008). Huhtanen (2008) added that complicated interactions between the animal and feed characteristics, and difficulties in distinguishing and quantifying these factors are one of many reasons. Huhtanen (2003) performed an in-depth data analysis with the goal of understanding the relationships between fermentation characteristics and milk production by comparing the dependencies between total acids and various milk components such as yield, fat, protein and ECM. This is an informative study on fermentation of the silage but does not translate to the overall quality of the silage in relation to milk production since fermentation is not the only Factor that defines the quality of silage. In addition, this study was conducted using grass silage and in reality, corn silage is as commonly used as grass silage. There are numerous examples of other studies that focus on a specific component of silage or milk variables. For examples, Kalac (2011) conducted research on ethanol and its relationship to milk production, and Havemose *et al. I2*004) studied the relationships between tocopherol and oxidative products of milk. In addition, Pang *et al*. (2019) focused their study on organic matter digestibility from different silages and how they affect milk production, while Kuoppala *et al.* 2008 conducted a study on how regrowth or primary silage can make differences in milk production. Individual analysis on specific components within silage or milk production does not validate true silage quality and therefore cannot be used to make decisive conclusions regarding the selection of silage type for optimal milk production. Tedeschi (2019) pointed out an interesting argument about how research in the past in agriculture and ruminant production lacks the adaptation of new technologies in the field of data analytics and artificial intelligence (AI) – such as machine learning – to fully take advantage of data-driven models. Tedeschi (2019) added that there is a need for novel data analytic methods that are developed for agriculture and ruminal production.

2.3 Application of Data Science in Dairy Science

With the rise of data-driven solutions, the application of data science in the field of dairy science is slowly gaining interest in academia. Big data is becoming more accessible to

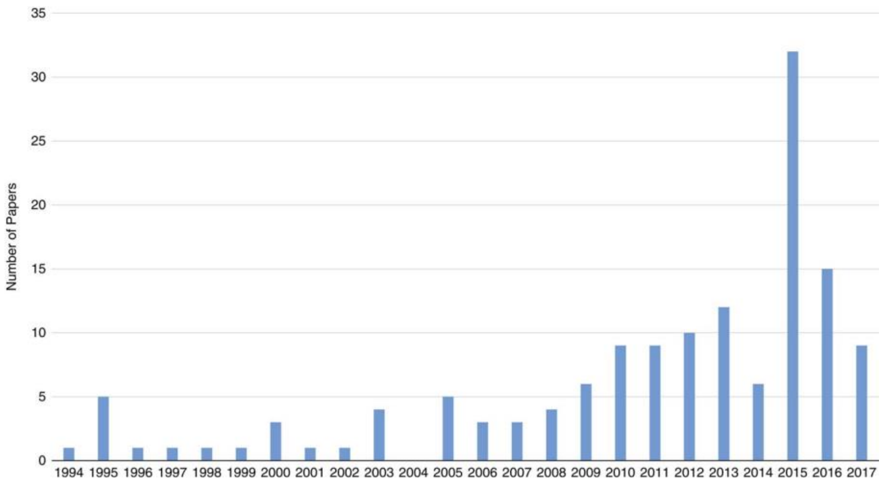


Figure 2.1: Number of published papers that use data science in dairy science (Lokhorst et al., 2019)

Table 2.2: Number of different machine learning approaches used in dairy science (Lokhorst et al., 2019)

Categories	Level	Number of papers: Level	Category
Supervised learning			134
	Regression	48	
	Classification	86	
Unsupervised learning			19
	Clustering	7	
	Dimensionality reduction	12	
Semi-supervised classification			1
Reinforcement learning			0
Total[1]			154

[1]The total exceeds the number of unique papers used in the review (n=142) since some papers cover more than one main category of level of Big Data analytics.

researchers due to the adaptation of laboratories and industry collaborators who are willing to provide large datasets to explore research interests in collaborative efforts. **Figure 2.1 and Table 2.2** illustrates the rise of published papers that focus on the application of data science to dairy science and the details of machine learning approaches used in these papers, respectively.

The international Precision Dairy Farming 2016 conference was a venue that presents numerous data-driven research (Kamphuis and Steeneveld, 2016). At that conference, Dias *et al.* (2016) addressed the creation of value with data from pasture-based farming systems, Van der Waaij *et al.* (2016) used machine learning to predict individual cow feed intake, Verhoosel and Spek (2016) examined the semantics for Big Data applications, Harty and Healy (2016) used Big Data advanced analytics to optimize health and fertility, and Bahr *et al.* (2016) discussed the field of data-driven smart feeding. During the European Conference on Precision Livestock Farming (Berckmans and Keita, 2017), there was an entire session specifically devoted to Big Data and its implication to the field of Data Science. The concept of data mining was used to find new insights and knowledge when data from several farms were brought together. Lokhorst *et al.* (1999) investigated the potential of data mining to benchmark dairy farms at the farm level. Although the development of new insights is also a promise of Big Data, the reviewed papers show a more structural analysis based on assumptions (and biological relevance) and supervised learning techniques. Currently, teams of multidisciplinary scientists at the university of Wisconsin-Madison led by Dr. Cabrera are very active in the field of data science and precision dairy farming. They developed an agricultural data hub called Dairy Brain which aims to perform various data science applications including nutritional grouping that provides a more accurate diet to lactating cows by automatically allocating cows to pens according to their nutritional requirements aggregating and analyzing data streams from management, feed, Dairy Herd Improvement (DHI), and milking parlor records (Cabrera *et al*., 2020). van der Heide *et al.* (2019) reported that using genetic data, prediction of survival by second lactation was possible and comparing statistical versus machine learning based models, machine learning models performed better. There was also a study on predicting milk production based on animal and dietary parameters using various machine learning techniques [Neural Network (NN), Random Forest (RF) and Support Vector Machine (SVM)], (Nguyen *et al.*, 2020). Gianola *et al.* (2011) conducted research on determining the relationship between wheat and milk production through Bayesian neural network (BNN).

2.4 Data Science

2.4.1 Defining a Data Science Problem

A Data Science problem is a problem that require data-driven solutions with the goal of understanding dependencies among data and predicting both known known and unknown outcome. Data science enables the transformation of big data into information, knowledge and action. Data science uses a combination of statistics and machine learning, which are also termed as statistical learning to gain insight and provide valuable solutions. Statistical techniques focus on determining the inference within datasets compared to machine learning, which focuses on prediction.

According to Lokhorst *et al.* (2019) supervised learning is a machine learning approach where the outcome of interest is known for each record used for model development. In other words, the data used for model development are labelled. Within this category, papers are classified into *regression* and *classification*. For regression, the outcome variable has a numerical value. Possible techniques involve linear regression, polynomial regression, use of radial basis functions, multivariate adaptive regression splines or multilinear interpolation. For classification, the outcome variable is categorical (e.g., binary yes/no). Possible techniques include neural networks, decision trees, naïve Bayes model or support vector machines.

Lokhorst *et al.* (2019) also defined unsupervised learning as a machine learning approach where the outcome of interest is unknown (unlabeled) for each record used for model development. Within this category, papers were classified on clustering and dimensionality reduction. Clustering techniques include K-means, Gaussian modelling, spectral or hierarchical clustering. Dimensionality-reduction techniques include, for example, principal component analysis (PCA), linear discriminant analysis (LDA) or independent component analysis.

2.4.2 Data Outliers

2.4.2.1 Univariate Outliers

Outlier detection, also known as anomaly detection, refers to the identification of statistically unlike observations which differ from the general distribution of datasets. It is also a term used to detect outliers that are not statistically defined but determined to be out of norm within a given domain. Usually, popular algorithms for outlier detection are for univariate outliers. However, univariate outlier detection is not ideal for datasets that have highly dependent variables when applying machine learning models. This is often the case for biochemical datasets since they are treated as a sample that consists of various dependent variables.

2.4.2.2 Multivariate Outliers

Python is becoming one of the most used programming languages for applying machine learning in the field of data science. However, there are no robust techniques to perform multivariate outlier detection (Zhao *et al.*, 2019). PyOD is an open-source Python toolbox for performing scalable outlier detection on multivariate data. This toolbox is rapidly gaining popularity in various fields of industries and research with interest in data science due to its state-of-the-art approach in multivariate outlier detection. Uniquely, it provides access to a wide range of outlier detection algorithms, including established outlier ensembles and more recent neural network-based approaches, under a single, well-documented Application

Table 2.3: List of Outlier Detection Techniques used in PyOD (Zhao et al., 2019)

Method	Category
LOF (Breunig et al., 2000)	Proximity
kNN (Ramaswamy et al., 2000)	Proximity
AvgkNN (Angiulli and Pizzuti, 2002)	Proximity
CBLOF (He et al., 2003)	Proximity
OCSVM (Schölkopf et al., 2001)	Linear Model
LOCI (Papadimitriou et al., 2003)	Proximity
PCA (Shyu et al., 2003)	Linear Model
MCD (Hardin and Rocke, 2004)	Linear Model
Feature Bagging (Lazarevic and Kumar, 2005)	Ensembling
ABOD (Kriegel et al., 2008)	Proximity
Isolation Forest (Liu et al., 2008)	Ensembling
HBOS (Goldstein and Dengel, 2012)	Proximity
SOS (Janssens et al., 2012)	Proximity
AutoEncoder (Sakurada and Yairi, 2014)	Neural Net
AOM (Aggarwal and Sathe, 2015)	Ensembling
MOA (Aggarwal and Sathe, 2015)	Ensembling
SO-GAAL (Liu et al., 2019)	Neural Net
MO-GAAL (Liu et al., 2019)	Neural Net
XGBOD (Zhao and Hryniewicki, 2018b)	Ensembling
LSCP (Zhao et al., 2019)	Ensembling

Programming Interface (API) designed for use by both practitioners and researchers. The key advantage of this toolbox is its capability of combining various outlier detection techniques listed in **Table 2.3** through a single API.

2.4.3 Data Transformation

2.4.3.1 General Information of Data Transformation

Data transformation is becoming the norm for data pre-processing. With the rise of machine learning, the focus is on achieving the highest-performance accuracy, and various studies have proven that data transformation can raise performance accuracy if applied appropriately. This is achieved by transforming datasets to come closer to suggested assumptions of models. However, data transformation must be applied with caution since they are not fit for any dataset and changing the original dataset can have critical consequences. More specifically, thorough research on finding the suitable data transformation technique must be performed to avoid misinterpretation of data.

Among the transformations employed in biological fields, the most used transformations are logarithmic, square root and angular (Ribeiro-Oliveira *et al*., 2018). These transformations are usually associated with non-normal data (Zar, 2014). Under such circumstances, data transformation is the most appropriate remedial measure (Dey, 2020). With the help of these techniques, the original data can be converted to a new scale resulting in a new dataset, which is expected to satisfy the variance homogeneity principle (Montgomery *et al*., 2017). Among listed techniques, logarithmic and angular are mostly used for data with percentages as units.

2.4.3.2 Logarithmic Transformation

The logarithmic transformation is suitable for cases such as when the variance is proportional to square of the mean, the coefficient of variation is constant or where effects are multiplicative (Montgomery *et al*., 2017; Dey *et al*., 2020). When the data range is wide, these conditions are usually found (Dey *et al*., 2020). This transformation is effective specifically in case of normalizing a positively skewed distribution (Dey *et al*., 2020). It is also helpful to achieve additivity (Zar, 2014; Rangaswamy, 2018). This transformation is commonly used for data with units of percentage, but it is not ideal for 0% and 100%.

2.4.3.3 Angular Transformation

Angular Transformation variables, expressed by a proportion or percentage, are best suited for the application of angular transformation (Zar, 2014) so that variance can be expressed as a quadratic function of the proportion (Warton *et al*., 2011). If the distribution of percentages is binomial (Dean, and Voss, 1999; Montgomery, 2013; Gupta, and Kapoor, 2014; Montgomery *et al*., 2017), this transformation makes the distribution closer to normal. It is also known as 'arcsine' or 'inverse sine' transformation. In agronomical experiments, biochemical components of silages are converted to proportions or percentages of total DM. It should be noted that only the percentage data that are derived from original data should be transformed (Rangaswamy, 2018).

2.4.4 Regression Models

2.4.4.1 Linear Models

A linear model is a sum of a constant and a product of parameter and predictor variable. It is possible that there are multiple products of parameter and predictor variable.

$$Y = b_o + b_1X_1 + b_2X_2 + ... + b_kX_k$$

Transforming the predictor variables in ways that produce curvature is also possible. For instance, you can include a squared variable to produce a U-shaped curve.

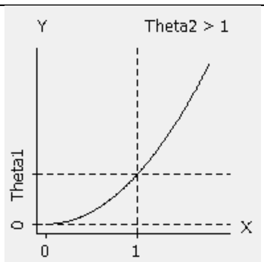
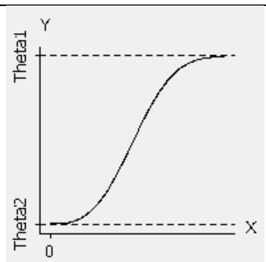
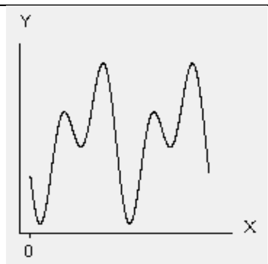
$$Y = b_o + b_1X_1 + b_2X_1^2$$

This model is still linear in the parameters even though the predictor variable is squared. You can also use log and inverse functional forms that are linear in the parameters to produce different types of curves.

2.4.4.2 Non - Linear Models

Non – Linear model can have more than one parameter per predictor variable. This allows flexibility of different curvatures to cover many shapes. **Table 2.4** presents some examples of non-linear functions and their possible shapes.

Table 2.4: Examples of Non-Linear Functions (The Minitab Blog, 2017)

Nonlinear function	Power (convex): $\mathrm{Theta1} \cdot \mathrm{X}^{\mathrm{Theta2}}$	Weibull growth: $\mathrm{Theta1} + (\mathrm{Theta2} - \mathrm{Theta1}) \cdot \mathrm{e}^{(-\mathrm{Theta3} \cdot \mathrm{X}^{\mathrm{Theta4}})}$	Fourier: Theta1 · cos(X + Theta4) + (Theta2 · cos(2X + Theta4)) + Theta3
One possible shape			

2.4.5 Multivariate Regression

Multivariate regression also known as multi-output, multi-target, multi-response regression aims to simultaneously predict multiple real-valued output/target variables (Borchani, 2015). When there are compound dependencies among feature/target variables, it has been proven that multivariate regression performs better than univariate regression (Borchani, 2015). The main advantage of multivariate regression is its capability to consider not only the underlying relationships among the features and the corresponding targets, but also the relationships among the targets, thereby guaranteeing a better representation and interpretability of real-world problems (Borchani, 2015). Other advantages include better computational efficiency and simpler models which tend to also have lower model training time. There are two common approaches to achieve multivariate regression.

First, the problem transformation methods (also known as local methods) transform the multi-output problem into independent single-output problems where each problem is solved using a single-output regression algorithm (Borchan, 2015). The key disadvantage of the problem transformation method is that it ignores the relationships among target variables.

Multi-Target Regressor Stacking (MTRS) is a popular problem transformation method that is commonly used. This method uses single output regression for each target variables and then builds another model based on the predicted target variables and corrects the errors made by the single output models. Regressor Chain (RC) is also a commonly used method that selects a random chain (order) of targets, and predicts targets based on the order of chain. The first model uses a single output regression to predict the first target variable and the following models will build a single output regression to predict the next target variable based on previous models which includes at least one target variable. Shortly after the development of this method, it evolved to Regressor Chain Corrected (RCC) and then Ensemble Regressor Chain Corrected (ERCC). The RCC is RC with cross validation and ERCC attempts to solve the problem of chain selection since depending on the order of the chain, predictions may vary. More specifically, ERCC with attempt every model if a combination of target variables results in distinct chains that are less than 10 and if they are more than 10, it will randomly choose 10 distinct chains (Spyromitros-Xioufis *et al*., 2012). Unfortunately, there are not enough studies to validate that MTR and RC improve predictive accuracies significantly compared to single output regression. Another study on Support Vector Machine (SVR) presented a Multi-Output SVR (MO-SVR) approach based on problem transformation (Zhang *et al*., 2012). It builds a multi-output model that considers the correlations among all the targets using the vector virtualization method (Zhang *et al*., 2012). Basically, it extends the original feature space and expresses the multi-output problem as an equivalent single-output problem, such that it can then be solved using the single output least squares SVR machines (LS-SVR) algorithm. It is at least as performant as a single output SVR and often faster in computations (Borchani, 2015).

Secondly, algorithm adaptation methods (also known as global or big-bang methods) adapt a specific single-output method (such as decision trees and support vector machines) to directly handle multi-output data sets (Borchani, 2015). Algorithm adaptation methods are deemed to be more challenging because they usually aim, not only to predict the multiple targets, but also to model and interpret the dependencies among those targets (Borchani, 2015). Statistical methods use correlations among target variables to make a build a single model that predicts all target variables. There are various statistical methods to perform the algorithm adaptation methods, but they commonly use a matrix of estimated regression coefficients. Various studies have also presented methods to use MO-SVR as part of algorithm adaptation methods by considering the correlations between target variables. To name a few, CoKriging method is a multi-output algorithm that exploits the correlations due to the proximity in the space of factors and outputs (Vasquez *et al*., 2003). In this way, with an appropriate choice of covariance and cross-covariances models, the authors showed that multi-output SVR yields better results than an independent prediction of the outputs (Vasquez *et al*., 2003). Other variant method is the multi-regressor SVR (M-SVR), which is based on an iterative reweighted least squares (IRWLS) procedure that iteratively estimates the weights W and the bias parameters b until convergence, i.e., until reaching a stationary point where there is no more improvement of the considered loss function (Sánchez-Fernández *et al*., 2004). Kernel methods use vector-valued learning where there is an emphasis on analyzing the regularized least squares from the computational point of view (Borchani, 2015). This method also analyzes the theoretical aspects of reproducing kernel Hilbert spaces (RKHS) in the range-space of the estimator and generalizing the representer theorem for Tikhonov regularization to the vector-valued setting (Borchani, 2015). Next, there is the multivariate regression trees (MRTs), also known as multi-objective regression trees (MORTs). This method not only identifies dependencies among target variables, but also builds a much smaller single regression tree. This method also inherits characteristics of univariate regression trees: they are easy to construct, and the resulting groups are often simple to interpret; they are robust to the addition of pure noise response and/or feature variables; they automatically detect the interactions among variables; and they handle missing values in feature variables with minimal loss of information (De'ath, 2002). Lastly, there is Rule methods, also known as Fitted Rule Ensemble (FIRE) algorithm, that transcribes an ensemble of regression trees into a large collection of rules, followed by a gradient-directed optimization procedure, to select the best (and much smaller) subset of these rules and determine their respective weights (Borchani, 2015).

2.4.6 Evaluation Metrics for Regression Problems

When evaluating regression models, there are four common evaluation metrics that are used. First, Mean Squared Error (MSE) is the most used regression model evaluation metric due to its simplicity.

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2$$

It is simply an average of the squared difference between target values and predicted values. As it squares the differences, it penalizes even a small error which leads to over-estimation of how bad the model is (Mishra 2019). It is preferred more than other metrics because it is differentiable and hence can be optimized better (Mishra, 2019).

Next, Root Mean Squared Error (RMSE) is the square root of the averaged squared difference between the target value and the predicted value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\mathrm{Predicted}_i - \mathrm{Actual}_i)^2}{N}}$$

It is preferred more in some cases because the errors are first squared before averaging which poses a high penalty on large errors (Mishra, 2019). This implies that RMSE is useful when large errors are undesired.

Mean Absolute Error (MAE) is the absolute difference between the target value and the predicted value.

$$MAE = \frac{1}{n} \sum | y - \hat{y} |$$

The MAE is more robust to outliers and does not penalize the errors as extremely as MSE. MAE is a linear score which means all the individual differences are weighted equally. It is not suitable for applications where you want to pay more attention to the outliers (Mishra, 2019).

Lastly, R-Squared also known as Coefficient of Determination or R^2 is another metric used for evaluating the performance of a regression model.

$$R^2 = 1 - \frac{\mathrm{MSE\ (model)}}{\mathrm{MSE\ (baseline)}}$$

These metrics help us to compare our current model with a constant baseline and tells us how much our model is better. The constant baseline is chosen by taking the mean of the data and drawing a line at the mean. R-Squared is a scale-free score that implies it does not matter whether the values are too large or too small, the R-Squared will always be less than or equal to 1. Sometime, R-Squared is a negative value and this is possible since the range of R-Squared is from negative infinity to 1 (Mishra, 2019). This is mainly due to the MSE of the model being greater than the MSE baseline. This is caused by multiple reasons which include the possibility of large amounts of outliers, a missing intercept, and a model that does not fit the trend of the data (Mishra, 2019).

2.5 Challenges in Understanding the Relationship Among Silage Qualities and Milk Production

Extensive studies on attempting to improve the interpretability of the impact of silage quality on milk production is necessary to facilitate industry leaders on optimizing efficiency of milk production. Data-driven analysis using data science approaches including machine learning and statistics will generate insightful in information on the relationship among silage characteristics and milk production variables.

3. Materials and Methods

The research was divided into two main components: i) silage data imputation; and ii) predictive modelling for silage factor analysis and its impact on milk production.

3.1 Silage Data Imputation

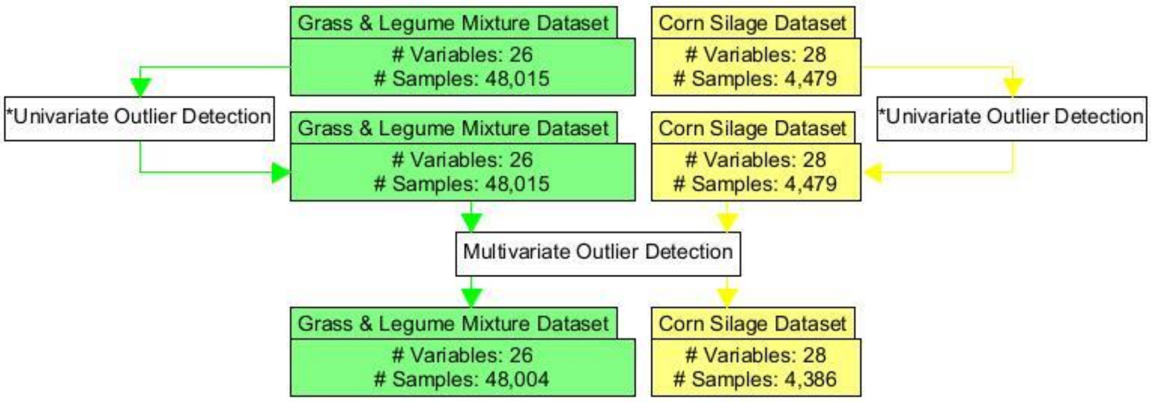
3.1.1 General Information about Datasets

The data used in this research originated in Québec, and were provided by Lactanet, the Canadian milk-recording agency. File sources contained information on silage samples from grass and legume, and corn silage; and milk production data (18,378, 34,237 and 47,785 observations, respectively). More specifically, the silage files consisted of chemical variables related to silage fermentation, digestibility, protein, fiber, carbohydrates and minerals, while the milk production file consisted of milk production and component variables from 2016 to 2020 (**Appendix 1**). Silage fed consists of a mixed ration of grass, legume and corn silage.

3.1.2 Data Editing

Data editing was performed using the Anaconda toolkit, which is an open-source Python distribution platform with numerous data science and machine learning packages, and SAS (Statistical Analysis System). Variables were selected by considering the Committee of Animal Nutrition from the National Research Council (NRC) and the common variables between silages and milk production datasets that could impact silage quality and milk production**.** There are 24 grass and legume silage and 26 corn silage variables and 37 milk production variables without the management related variables (**Appendix 1**).

As an initial step of data correction, data were checked for duplicates as well as herds or years with a single sample. Since all variables except pH are in the unit of percentage, data with values less than 0 or greater than 100 were eliminated. In addition, pH values less than 0 and greater than 14 were also eliminated. Other outliers were determined and eliminated as follows: for grass and legume silage, there were three values greater than 100 percent, one for acid detergent fiber with neutral detergent fiber percentage as the unit and two for neutral detergent insoluble crude protein with crude protein percentage as the units. For corn silage, two values were greater than 100 percent, one for rumen digestible neutral detergent fiber at 120 hours *in vitro* and one for rumen digestible Neutral Detergent Fiber at 240 hours *in* *vitro*. Rules for biochemical outliers were set differently for the two silage types. For corn silage dataset, starch values less than 12 percent and greater than 47 percent, dry matter values less than 25 percent and greater than 54 percent and crude protein values greater than 15 percent were considered outliers. Since there are dependencies among these variables, multivariate outlier detection, by looking at the entire sample, was also necessary. Isolation Forest algorithm is a multivariate outlier technique that uses a set of trees to perform data partitioning it provides an anomaly score looking at how isolated a point is in the structure. The anomaly score is then used to identify outlier samples from normal observations. Details of the changes in the number of variables and samples are shown in **Figure 3.1.**



*Note that the number of samples does not change after the univariate outlier detection since it only eliminates data values, not the entire sample. The number of variables only changes during variable selection, where only the variables that will be used in the project are selected.

Figure 3.1: Variable Selection and Outlier Detection for Silage Datasets

3.1.3 Data Imputation

In order to study the relationships among the silage and milk production variables, these variables needed to be present in the same sample. It is not common in forage and dairy industries to have forage dataset and milk production dataset in a single database. The milk-production dataset used for this research had both milk production and silage variables, but some important silage variables listed as target variables in **Appendix 2** were not available. Therefore, it was necessary to predict numerical values of the target variables using the silage variables provided within the milk production dataset as an input to a machine learning model that was built using the training variables listed in **Appendix 2**.

Throughout this research, various machine learning algorithms were explored to select the most performant regression algorithm. Since the datasets were all labelled with known silage and milk production variables, supervised learning used algorithms were explored. The imputation is a multi-input multi-output regression task since there are multiple variables to predict, and both the single-output and multi-output regression approaches were considered.

Prediction, using multi-output approaches such as multi-regressor and regressor chain, were also explored. As a result, Extra Tree (Extremely Randomized Trees)-based Meta-estimator and regressor chain based AdaBoost were the best algorithms for grass and legume silage and corn silage datasets, respectively. For grass and legume silage dataset, 969 samples were used to train the model and to impute 21,031 samples. For the corn-silage dataset, 2,927 samples were used to train the model and to impute 12,818 samples. Details of the data imputation process can be shown in **Figure 3.2**. Initially, linear models such as linear regression, robust (Huber) regression, partial least square (PLS) regression, Elastic-Net regression and k-nearest neighbour regression were explored. Mean Absolute Error (MAE), MAE/Mean, Root Mean Square Error (RMSE) and R-square were all used to evaluate the accuracy of the models. Data normality was tested by observing the histograms of each variable and in general, silage variables did not show normal distribution. To bring the data closer to normality, we explored numerous data transformation techniques and settled with the arcsine square root transformation. This data transformation technique is for percentage data and was, therefore, applied to all variables except pH. The application of data transformation made small changes to the distribution of the data points, but slightly improved the R-Square at the expense of a slight decrease in MAE.

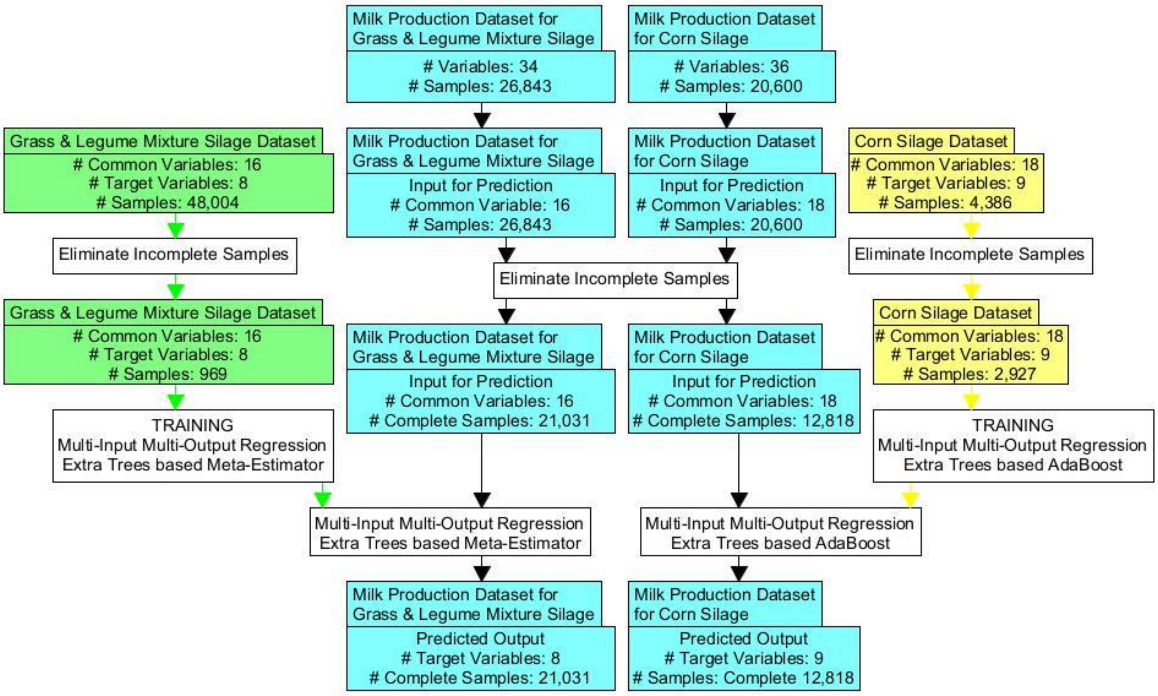


Figure 3.2: Data Imputation Process Through the use of Machine Learning Algorithms

3.2 Predictive Modelling for Silage Quality Factor Analysis and its Impact on Milk Production

3.2.1 Factor Analysis for Silage Quality

Using the imputed silage variables, factor analysis was conducted to provide better interpretability and understand the impact of silage characteristics on milk production. By using the factor loadings (coefficients that explain the correlation between a factor and a variable) derived from Varimax rotation, factor scores were calculated and by doing a dot product between the factor loadings and silage values.

Factors for grass and legume silage are denoted as Fx_{GL}, where x represents 1 to 5. To facilitate interpretability, some factor variables needed their signs to be inverted by multiplying them by negative one to represent higher values as positive outcomes. In detail, $F1_{GL}$ represents neutral detergent fiber digestibility and the Factor loadings of neutral detergent fiber digestibility variables had negative values, which were converted to positive values to designate higher as better. $F5_{GL}$, which represents soil contamination, had positive ash, which was converted to negative since lower the ash, less the soil contamination. In order to have a truly complete dataset, additional multivariate outliers were required. Herds that had a single sample for an entire year were removed.

Factors for corn silage are denoted as $Fx_{C,}$ where x represents 1 to 5. To facilitate interpretability, some factor variables needed their signs to be inverted by multiplying them by negative one to represent higher values as positive outcomes. In detail, $F1_{C}$, which represents starch concentration, was converted from negative to positive values to emphasize higher starch concentration. In order to have a truly complete dataset, additional multivariate outliers were required. Herds that had a single sample for an entire year were removed.

3.2.2 Linear Models

Linear regression was used to understand the relationships between milk management variables and silage quality factors. More specifically, each factor was compared to test month and storage method, where the test month comes from the herd test period which is the milk recording date and not the month of the harvesting or the analysis of the sample.

3.2.3 Mixed Effects Models

Mixed effects models were used to understand the relationships among the silage characteristics, herd management and milk production variables. More specifically, five models were built to represent average daily milk, fat, protein milk urea nitrogen and somatic cell count score per cow. Input parameters are defined in **Table 3.1**.

Table 3.1: Information about the Selected Variables for Mixed Effects Model

VARIABLE	FIXED EFFECTS	RANDOM EFFECTS	OUTPUT	UNITS
SumPropFactorGL 1	X			N / A
SumPropFactorGL 2	X			N / A
SumPropFactorGL 3	X			N / A
SumPropFactorGL 4	X			N / A
SumPropFactorGL 5	X			N / A
SumPropFactorC 1	X			N / A
SumPropFactorC 2	X			N / A
SumPropFactorC 3	X			N / A
SumPropFactorC 4	X			N / A
SumPropFactorC 5	X			N / A
Average Daily Milk per Cow			X	kg / day / cow
Average Daily Fat per Cow			X	kg / day / cow
Average Daily Protein per Cow			X	kg / day / cow
Somatic Cell Count Score			X	Cell count average
Average Days in Milk			X	Number of days
Test Year	X			Year
Test Month	X			Month
Herd		X		ID
Region	X			Region
Proportion of Concentrate	X			%
Milk Urea Nitrogen Average	X			mg/dl
Presence of Other Silage	X			%

N / A: Not Available
SumPropFactorGL: Sum of the Interaction between Proportion of Grass and Legume Silage within Feed and Grass and Legume Silage Factor
SumPropFactorC: Sum of the Interaction between Proportion of Corn Silage within Feed and Corn Silage Factor

Five variables were created for each grass and legume silage and corn silage that represent the herd level sum of the interactions between the proportion of the specific silage (grass and legume silage or corn) and the factors. More specifically, each factor score was multiplied by the proportion of its silage type within feed and summed within a herd. Grass and legume silage and corn silage had different models due to their different silage quality factors.

The mixed effects models were formulated by the following equation:

$$Y = \sum_{x=1}^{5} \sum_{\text{sample}}^{TotalSampleHTP} \frac{\text{Proportion of specific silage}}{\text{from feed (DM\%)}} \cdot \text{Factor}_{\text{x}} +$$

Days in Milk + Test Year + Test Month + Region + Proportion of Concentrate from Feed (DM %) + Total Proportion of Other Silage from feed (DM %) + Random (Herd)

TotalSampleHTP: Total Sample Count Within Herd Test Period

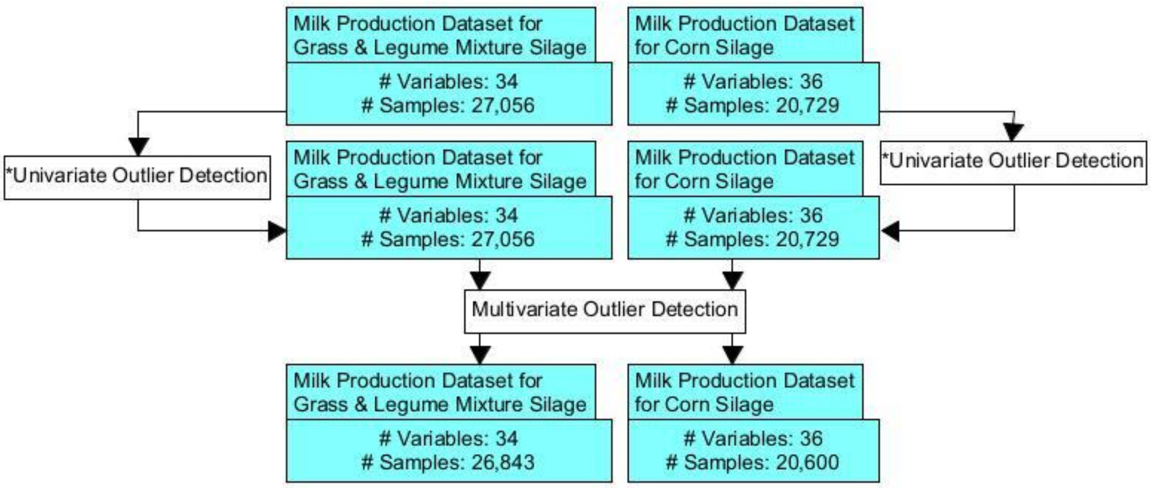
Where Y is the following:

- Average daily milk per cow (kg);
- Average daily fat per cow (kg);

- Average daily protein per cow (kg);
- Average daily milk urea nitrogen;
- Average somatic cell score

Specifically, silage is either predominantly grass and legume silage or predominantly corn silage. However, the proportion of the other silage is also included in the statistical model. Herd is a random effect.

Additional univariate outlier removal was necessary before applying the mixed effects model. In order to have a complete dataset, an entire set of samples that belong to the same herd test period had to be removed if a value was either missing or an outlier: proportion of total grass and legume silage less than 10%, proportion of total corn silage greater than 70%, average days in milk greater than 260, average daily milk per cow daily in a herd less than 15kg, average milk urea nitrogen that equal to 0, average somatic cell count that equals to 0 or greater than 800, proportion of concentrate in feed greater than 60%, proportion of total silage from total feed less than 30%, total daily silage intake per cow less than 10kg, average daily dry matter intake per cow less than 15 kg or greater than 35 kg and total milk produced from concentrate (kg) less than 2 or greater than 7 were considered univariate outliers in milk production dataset (**Figure 3.3**).



*Note that the number of samples does not change after the univariate outlier detection since it only eliminates data values, not the entire sample. The number of variables only changes during variable selection, where only the variables that will be used in the project are selected.

Figure 3.3: Variable Selection and Outlier Detection for Milk Production Datasets

4. Results and Discussion

4.1 Interpretation of Grass and Legume Silage Quality Imputation, Factors and Relationship with Milk Production Management

4.1.1 Silage Quality Imputation Results

Within the milk production dataset, 5,812 grass and legume silage samples had missing input parameters and therefore did not impute the rest of the silage variables. As a result, final dataset includes 450 herds and 5,539 samples for the milk production dataset with mainly grass and legume silage (**Figure 4.1**). Various regression algorithms were explored to identify the algorithm with ideal prediction accuracy of grass and legume silage qualities. Despite having some algorithms perform better for silage particular variables with particular metrics, an algorithm that generally performs well for all silage variables was chosen for efficiency. Highest accuracy for data imputation was found using ensemble algorithms from supervised machine-learning with a performance of approximately 93 percent (**Table 4.1**). More specifically, Extra Tree (Extremely Randomized Trees)-based Meta-estimator was the best algorithm for grass and legume silage. The imputed silage dataset projected expected silage characteristics similar to the training dataset.

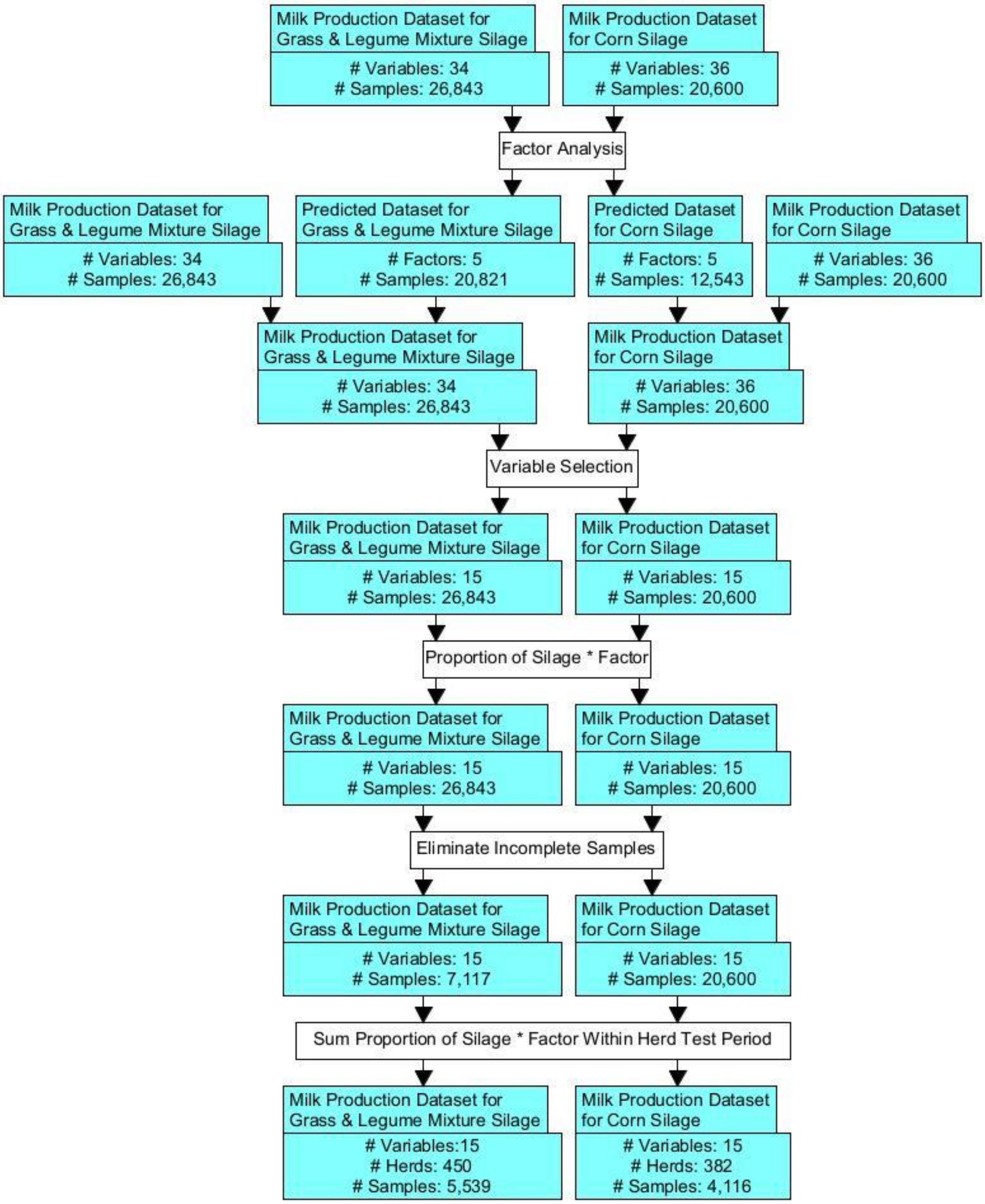


Figure 4.1: Data Preparation for the Mixed Effects Model

The accuracy metrics for the simple linear models were not high, especially the R-square for Butyric acid of grass and legume silage dataset (**Table 4.1**). Various feature selection techniques were explored, but they did not improve the R-Square. Data transformation and ensemble models performed significantly better as shown in **Table 4.1**.

Table 4.1: Machine Learning Model Prediction Metrics for Grass and Legume Silage Dataset

Variable	Model	Metric				
		Mean	MAE[1]	MAE / Mean	RMSE[2]	R^2
Rumen Digestible Neutral Detergent Fiber at 30 Hour in Vitro	Linear Regression	0.84	0.02	0.02	0.03	0.67
	Robust (Huber) Regression	0.84	0.02	0.02	0.02	0.89
	PLS (Partial Least Square) Regression	62.13	0.02	0.02	0.03	0.85
	ElasticNet Regression	0.84	0.02	0.02	0.02	0.89
	KNN (K Nearest Neighbour)	0.84	0.004	**0.01**	0.01	**0.96**
	SVR (Support Vector Regression)	62.13	1.93	0.03	2.76	0.83
	Meta (Bagging) Estimator	62.13	0.91	**0.01**	1.62	0.94
Rumen Digestible Neutral Detergent Fiber at 120 Hour in Vitro	Linear Regression	0.91	0.03	0.03	0.04	0.66
	Robust (Huber) Regression	0.91	0.02	0.03	0.03	0.85
	PLS (Partial Least Square) Regression	70.96	0.03	0.03	0.04	0.82
	ElasticNet Regression	0.91	0.02	0.03	0.03	0.86
	KNN (K Nearest Neighbour)	0.91	0.004	**0.01**	0.01	**0.96**
	SVR (Support Vector Regression)	70.96	2.72	0.03	3.91	0.76
	Meta (Bagging) Estimator	70.96	1.16	**0.01**	2.32	0.90
Rumen Digestible Neutral Detergent Fiber at 240 Hour in Vitro	Linear Regression	0.95	0.03	0.03	0.04	0.66
	Robust (Huber) Regression	0.95	0.02	0.02	0.03	0.87
	PLS (Partial Least Square) Regression	74.88	0.03	0.03	0.04	0.83
	ElasticNet Regression	0.95	0.02	0.02	0.03	0.87
	KNN (K Nearest Neighbour)	0.95	0.004	**0.01**	0.01	**0.96**
	SVR (Support Vector Regression)	74.88	2.44	0.03	3.56	0.79
	Meta (Bagging) Estimator	74.88	1.14	**0.01**	2.33	0.90
pH	Linear Regression	4.64	0.07	**0.01**	0.11	0.57
	Robust (Huber) Regression	4.64	0.17	0.03	0.23	0.67
	PLS (Partial Least Square) Regression	4.75	0.18	0.03	0.24	0.65
	ElasticNet Regression	4.64	0.17	0.03	0.23	0.67
	KNN (K Nearest Neighbour)	4.64	0.004	**0.01**	0.01	0.11
	SVR (Support Vector Regression)	4.75	0.16	0.03	0.21	0.68
	Meta (Bagging) Estimator	4.75	0.06	**0.01**	0.1	**0.94**
Lactic Acid	Linear Regression	0.20	0.02	0.13	0.03	0.70
	Robust (Huber) Regression	0.20	0.02	0.10	0.02	0.79
	PLS (Partial Least Square) Regression	3.83	0.02	0.11	0.03	0.75
	ElasticNet Regression	0.20	0.02	0.11	0.02	0.78
	KNN (K Nearest Neighbour)	0.20	0.00	0.22	0.01	0.66
	SVR (Support Vector Regression)	3.83	0.75	0.19	1.01	0.77
	Meta (Bagging) Estimator	3.83	0.29	**0.07**	0.50	**0.95**
Acetic Acid	Linear Regression	0.15	0.02	0.14	0.02	0.50
	Robust (Huber) Regression	0.15	0.02	0.13	0.02	0.83
	PLS (Partial Least Square) Regression	1.98	0.02	0.16	0.03	0.78
	ElasticNet Regression	0.15	0.02	0.14	0.02	0.82
	KNN (K Nearest Neighbour)	0.15	0.004	0.21	0.01	0.52
	SVR (Support Vector Regression)	1.98	0.63	0.31	0.88	0.57
	Meta (Bagging) Estimator	1.98	0.20	**0.10**	0.32	0.93
Propionic Acid	Linear Regression	0.01	0.01	0.18	0.01	0.47
	Robust (Huber) Regression	0.01	0.01	0.86	0.01	0.42
	PLS (Partial Least Square) Regression	0.33	0.01	1.15	0.01	0.40
	ElasticNet Regression	0.01	0.01	0.98	0.01	0.50
	KNN (K Nearest Neighbour)	0.01	0.004	**0.06**	0.01	0.32
	SVR (Support Vector Regression)	0.33	0.06	0.18	0.08	0.82
	Meta (Bagging) Estimator	0.33	0.02	**0.06**	0.03	**0.96**
Butyric Acid	Linear Regression	0.04	0.03	0.72	0.03	0.35
	Robust (Huber) Regression	0.04	0.02	0.71	0.03	0.32
	PLS (Partial Least Square) Regression	0.70	0.03	0.80	0.04	0.29
	ElasticNet Regression	0.04	0.01	0.98	0.01	0.50
	KNN (K Nearest Neighbour)	0.04	0.004	**0.04**	0.01	0.32
	SVR (Support Vector Regression)	0.70	0.18	0.26	0.34	0.19
	Meta (Bagging) Estimator	0.70	0.03	0.75	0.03	**0.96**

[1]MSE: Mean Absolute Error

[2]RMSE: Root Mean Squared Error

4.1.2 Silage Quality Factor Analysis

Grass and legume silage variables were grouped into five factors:

- $F1_{GL}$: Neutral Detergent Fiber Digestibility and Heat Damage
- $F2_{GL}$: Protein Ruminal Degradability
- $F3_{GL}$: Legume Proportion
- $F4_{GL}$: Homolactic Fermentation
- $F5_{GL}$: Soil Contamination and Bad Fermentation Pattern

In order to understand the relationship among the silage characteristics and milk production, the initial step was to analyze the impact of test date, region and storage method on factors derived from the factor loadings (**Table 4.2**). JMP statistical software was used to perform statistical analysis. For each factor, Tukey results with least square means were generated to observe statistical differences among the variables.

Table 4.2: Factor Loadings for Grass and Legume Silage

Factor and Variable	$F1_{GL}$	$F2_{GL}$	$F3_{GL}$	$F4_{GL}$	$F5_{GL}$	Communalities
Dry Matter (%)	15.15	-39.07	21.40	-66.98	24.38	0.73
Crude Protein (DM %)	5.67	-4.47	90.22	6.50	-25.63	0.89
Soluble Protein (DM %)	-0.35	61.85	53.67	38.48	-24.89	0.88
Soluble Protein (CP %)	-4.98	82.34	-1.76	41.74	-13.60	0.87
Acid Detergent Insoluble Crude Protein (DM %)	-42.32	-78.76	16.17	-0.25	-9.43	0.83
Acid Detergent Insoluble Crude Protein (CP %)	-43.60	-73.28	-29.86	-4.29	5.38	0.82
Neutral Detergent Insoluble Crude Protein (DM %)	-1.08	-90.02	19.51	-11.39	-1.57	0.86
Neutral Detergent Insoluble Crude Protein (CP %)	-4.71	-89.43	-26.88	-14.63	13.99	0.92
Rumen Degradable Protein (DM %)	3.15	28.99	81.75	23.79	-28.25	0.89
Rumen Degradable Protein (CP %)	-4.95	82.33	-1.73	41.75	-13.55	0.87
Acid Detergent Fiber (DM %)	-40.00	9.71	-78.72	13.98	-37.67	0.95
Acid Detergent Fiber (NDF %)	-79.44	23.75	9.64	17.75	-33.16	0.84
Neutral Detergent Fiber (DM %)	5.21	-5.12	-91.57	3.46	-19.21	0.88
Lignin (DM %)	-86.80	-39.40	-15.11	-0.77	-12.67	0.95
Lignin (NDF %)	-86.27	-32.08	30.96	-5.19	-3.29	0.95
Crude Fat (DM %)	38.43	3.90	15.65	58.56	-17.47	0.55
Non-fibrous Carbohydrate (DM %)	-7.46	-17.37	63.42	-10.91	55.30	0.76
Ash (DM %)	-14.23	18.65	17.11	-10.07	-44.27	0.29
Rumen Digestible Neutral Detergent Fiber at 30 Hours in Vitro (NDF %)	90.29	1.43	19.17	4.02	-5.05	0.86
Rumen Digestible Neutral Detergent Fiber at 120 Hours in Vitro (NDF %)	93.50	1.71	7.86	0.26	-1.37	0.88
Rumen Digestible Neutral Detergent Fiber at 240 Hours in Vitro (NDF %)	94.49	2.67	5.46	2.00	-0.37	0.90
pH	-2.59	-18.00	2.37	-89.03	-41.13	1.00
Lactic Acid (DM %)	-0.91	42.19	11.62	73.88	17.16	0.77
Acetic Acid (DM %)	-27.36	11.73	10.15	43.75	-41.11	0.46
Propionic Acid (DM %)	-11.74	40.06	-40.21	41.60	-59.71	0.87
Butyric Acid (DM %)	3.84	-7.32	-4.77	-4.88	-54.89	0.31
Eigenvalues	7.78	5.23	4.31	2.60	1.97	N / A
Explained Variance	22.14	19.74	14.39	9.74	8.58	N / A

$F1_{GL}$: Neutral Detergent Fiber Digestibility and Heat Damage
$F2_{GL}$: Protein Ruminal Degradability
$F3_{GL}$: Legume Proportion
$F4_{GL}$: Homolactic Fermentation
$F5_{GL}$: Soil Contamination and Bad Fermentation Pattern
N / A: Not Applicable
Green and red colors represent positive and negative values, respectively.

Table 4.3: Grass and Legume Silage Variables Used in Mixed Effects Modelling and Their Effects

Dependent Variables	Independent Variables	Estimates (+, -)	p-values
Average Daily Milk per Cow (kg)	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	-0.05	< 0.05
	Proportion of Concentrate from Feed	2.99	< 0.05
	Total Proportion of Corn Silage	1.50	< 0.05
	$F1_{GL}$	0.09	< 0.05
	$F2_{GL}$	0.13	< 0.05
	$F3_{GL}$	0.13	0.81
	$F4_{GL}$	-0.06	0.13
	$F5_{GL}$	-0.04	0.60
Average Daily Fat per Cow (kg)	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	-0.001	< 0.05
	Proportion of Concentrate from Feed	0.11	< 0.05
	Total Proportion of Corn Silage	0.05	< 0.05
	$F1_{GL}$	0.004	< 0.05
	$F2_{GL}$	0.006	< 0.05
	$F3_{GL}$	-0.01	0.92
	$F4_{GL}$	-0.35	0.10
	$F5_{GL}$	-0.07	0.81
Average Daily Protein per Cow (kg)	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	-0.001	< 0.05
	Proportion of Concentrate from Feed	0.10	< 0.05
	Total Proportion of Corn Silage	0.08	< 0.05
	$F1_{GL}$	0.003	< 0.05
	$F2_{GL}$	0.003	0.21
	$F3_{GL}$	-0.0006	0.95
	$F4_{GL}$	-0.001	0.31
	$F5_{GL}$	-0.007	0.78
Average Somatic Cell Score	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	0.001	< 0.05
	Proportion of Concentrate from Feed	0.08	0.50
	Total Proportion of Corn Silage	0.41	< 0.05
	$F1_{GL}$	-0.001	0.32
	$F2_{GL}$	-0.04	0.84
	$F3_{GL}$	-0.009	0.24
	$F4_{GL}$	0.005	0.47
	$F5_{GL}$	0.001	0.62
Average Daily Milk Urea Nitrogen	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	-0.001	0.21
	Proportion of Concentrate from Feed	0.09	0.83
	Total Proportion of Corn Silage	1.87	< 0.05
	$F1_{GL}$	-0.07	< 0.05
	$F2_{GL}$	-0.05	0.38
	$F3_{GL}$	-0.04	0.12
	$F4_{GL}$	0.09	< 0.05
	$F5_{GL}$	0.13	0.11

$F1_{GL}$: Neutral Detergent Fiber Digestibility and Heat Damage
$F2_{GL}$: Protein Ruminal Degradability
$F3_{GL}$: Legume Proportion
$F4_{GL}$: Homolactic Fermentation
$F5_{GL}$: Soil Contamination and Bad Fermentation Pattern
N/A: Not Available.
Estimates are highlighted in green for positive values and red for negative values.

4.1.2.1 Impact of Silage Management Variables on Neutral Detergent Fiber Digestibility and Heat Damage with Grass and Legume Silage

The factor $F1_{GL}$ (**Table 4.2**) was defined as an index related to fiber digestibility and heat damage because the factor loadings (**Table 4.2**) for rumen digestible neutral detergent fiber digestibility (NDFD) in vitro (30, 120, 240 hours) were 90.29, 93.50 and 94.49 respectively. The factor loadings for lignin (%DM and %NDF), ADF (%DM and %NDF) and CP bound to ADF (%DM and %CP) were the opposite of NDFD and significantly negative. The concentration of CP bound to ADF (Yu, 1976) was recognized as an indicator of heat-damaged silage during feed-out.

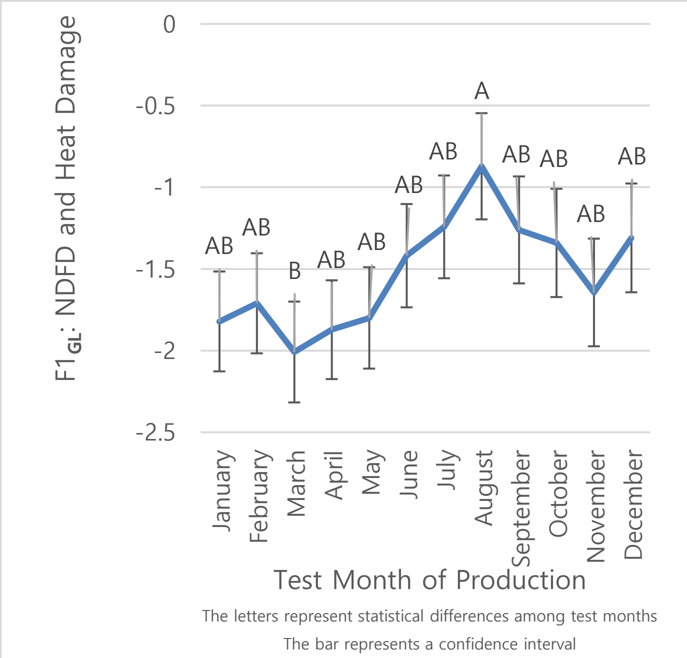


Figure 4.2: Effect of Test Month of Production for $F1_{GL}$

The results of the mixed model analysis for $F1_{GL}$ (**Table 4.3**) confirm the interpretation of the factor. A significative difference was observed between the test month of August (-0.87) and March (-2.01) (**Figure 4.2**). The difference is due to the higher risk of heat damage during feed-out in summer months compared to the winter month like observed by Bernardes *et al.* (2018). Also, the effect of storage methods on $F1_{GL}$, showed that the round and large rectangular bales (**Figure 4.3**) had a value of 0.30 and -0.34, respectively,

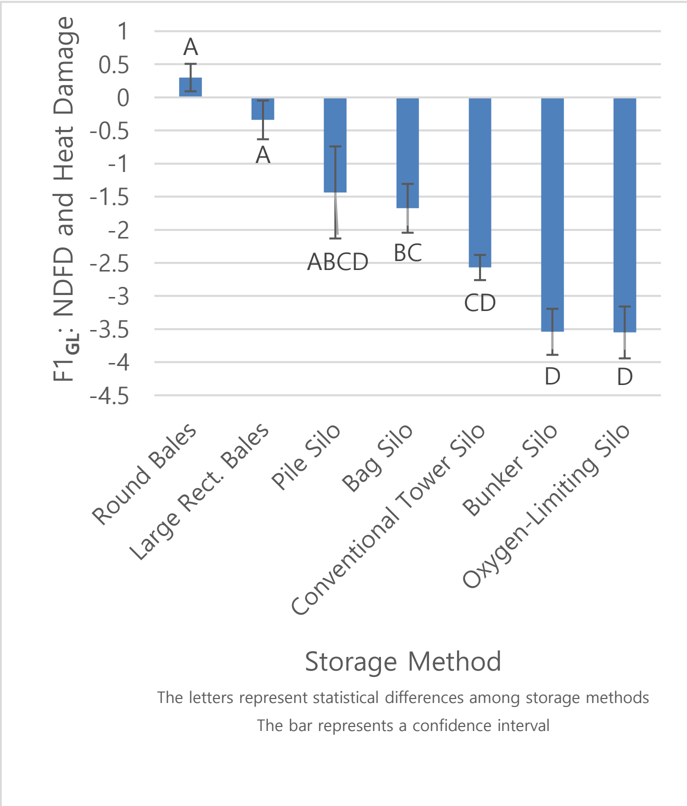


Figure 4.3: Effect of Storage Method for $F1_{GL}$

compared to bunker silos and the oxygen-limiting silos with -3.54 and -3.55, respectively. Pile silos, bag silos and conventional tower silos have intermediated results. Similar results have been found where bales exhibit reduced heat damage compared to other storage methods (Corbett, 1996; Coblentz *et al.,* 2009). The lower value for the oxygen-limited silos could be due to the improper management of this type of silos where a limited number of Quebec producers use dry ice for closing the silo or injecting CO_2 during the summer (Leduc, 2019) leading to heat damage.

4.1.2.2 Impact of Silage Management Variables on Protein Ruminal Degradability with Grass and Legume Silage

The factor $F2_{GL}$ (**Table 4.2**) was defined as an index related to protein ruminal degradability because the factor loadings (**Table 4.2**) for rumen degradable protein (%CP) and soluble protein (%DM and %CP) were 82.33, 61.85 and 82.34 respectively. Soluble protein fraction represents the portion of crude protein that is rapidly degraded or digested by rumen microbes. The factor loadings for ADICP (%DM and %CP) and NDICP (%DM and %CP) were the opposite of rumen degradable and soluble protein, and significantly negative. ADICP escapes ruminal breakdown and represents the portion of the protein that is not degradable, and NDICP is slowly degraded in the rumen and constitutes a major portion of the ruminal undegraded protein content (Sniffen *et al.*, 1992; NRC 1996).

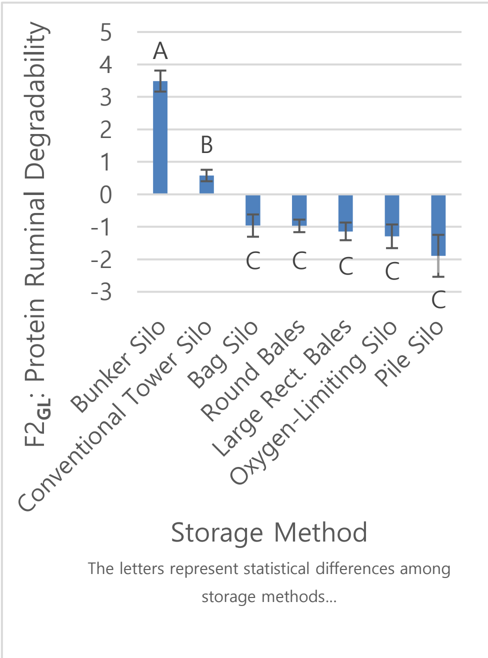


Figure 4.4: Effect of Storage Method for $F2_{GL}$

The results of the mixed model analysis for $F2_{GL}$ (**Table 4.3**) confirm the interpretation of the factor. Test month of production for $F2_{GL}$ (protein ruminal degradability) had no statistical differences among them, although it is known that protein degradability decreases before winter due to the increase of plant maturity (Hoffman *et al.*, 1993). The effect of storage methods on $F2_{GL}$, showed that the bunker silo (**Figure 4.4**) had a value of 3.49, compared to conventional tower silo with 0.58. Additionally, round bales, large rectangle bales, oxygen-limiting silo and pile silo had a value of -0.97, -1.14, -1.29 and -1.89, respectively.

4.1.2.3 Impact of Silage Management Variables on Legume Proportion with Grass and Legume Silage

The factor $F3_{GL}$ (**Table 4.2**) was defined as an index related to legume proportion because the factor loadings (**Table 4.2**) for rumen degradable, crude and soluble protein (%DM) were 81.75, 90.22 and 53.67 respectively. The factor loadings for ADF (%DM) and NDF (%DM) were the opposite of rumen degradable, crude and soluble protein, and significantly negative. The ADF and NDF (Telleng *et al.*, 2017) were recognized as an indicator of legume proportion within silage. The factor loading for NFC (%DM) was 63.42 and Villalba *et al.* (2021) observed NFC as another indicator of legume proportion within silage.

The results of the mixed model analysis for $F3_{GL}$ (**Table 4.3**) confirm the interpretation of the factor. Test month of production and $F3_{GL}$ (legume proportion) showed no statistical difference. This is likely due to having evenly distributed legume proportion as part of mixture silage throughout the year in Québec. It is common to observe legume as part of grass silage to enhance crude protein amount.

4.1.2.4 Impact of Silage Management Variables on Homolactic Fermentation with Grass and Legume Silage

The factor $F4_{GL}$ (**Table 4.2**) was defined as an index related to homolactic fermentation because the factor loadings (**Table 4.2**) for lactic, acetic and propionic acids were 73.88, 43.75 and 41.60 respectively. The factor loadings for DM and pH were the opposite of lactic, acetic and propionic acids, and significantly negative. Factor loading (**Table 4.2**) of crude fat was 58.56 and it (Solorzano *et al.*, 2016) was recognized as an indicator of homolactic fermentation.

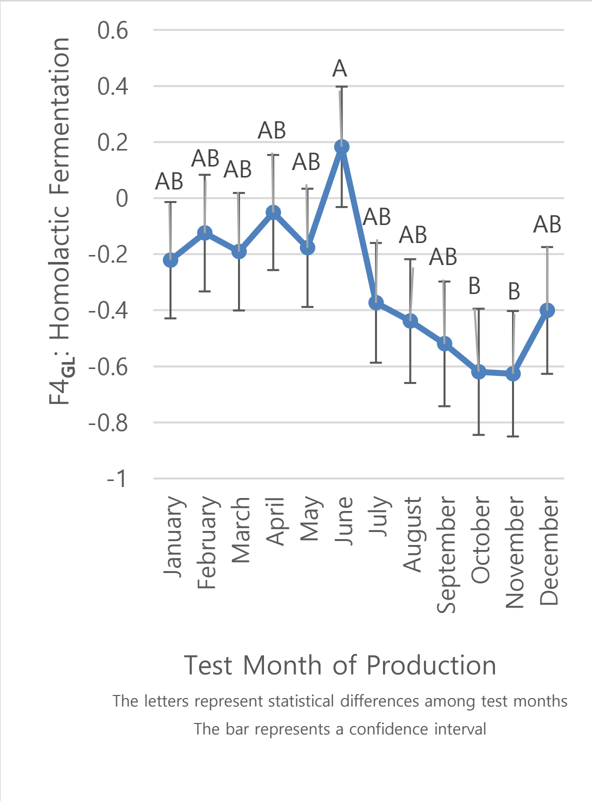


Figure 4.6: Effect of Test Month of Production for $F4_{GL}$

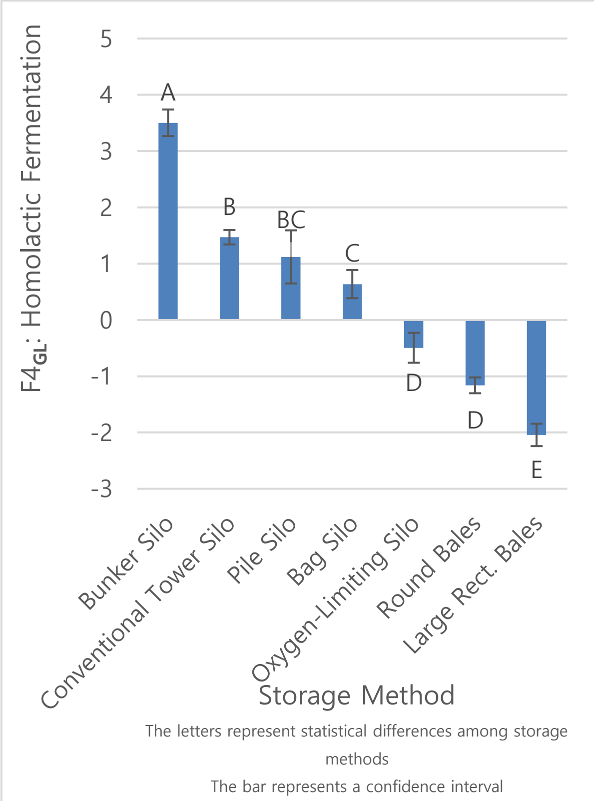


Figure 4.5: Effect of Storage Method for $F4_{GL}$

The results of the mixed model analysis for $F4_{GL}$ (**Table 4.3**) confirm the interpretation of the factor. A significative difference was observed in June (0.18) compared to September (-0.52) and October (-0.62) (**Figure 4.5**). The difference is due to the higher risk of heat damage during feed-out in summer months compared to the winter month like observed by Bernardes *et al.* (2018). Also, the effect of storage methods on $F4_{GL}$, showed that oxygen-limiting silos and round bales (**Figure 4.6**) had a value of -0.50 and -1.16, respectively, compared to other storage methods such as bunker silos, conventional tower silos, bag silos and large rectangle bales with 3.50, 1.47, 0.64 and -2.04, respectively. Pile silos has intermediated results. These results are likely due to bales having slower fermentation caused by being loosely packaged and having more oxygen (Schick, 2019). The higher values for bunker and conventional tower silo are likely due to the longer fermentation time on the two methods (Oney *et al.,* 2018).

4.1.2.5 Impact of Silage Management Variables on Soil Contamination and Bad Fermentation Pattern with Grass and Legume Silage

The factor $F5_{GL}$ (**Table 4.2**) was defined as an index related to soil contamination because the factor loading (**Table 4.2**) for ash was -44.27. The factor loading for NFC was the opposite of ash and significantly negative. Low pH (-41.13), acetic acid (-41.11), propionic acid (-59.71) and butyric acid (-54.89) were observed (**Table 4.2**), and these variables (Danner *et al.,* 2003) were recognized as an indicator of bad fermentation in silage.

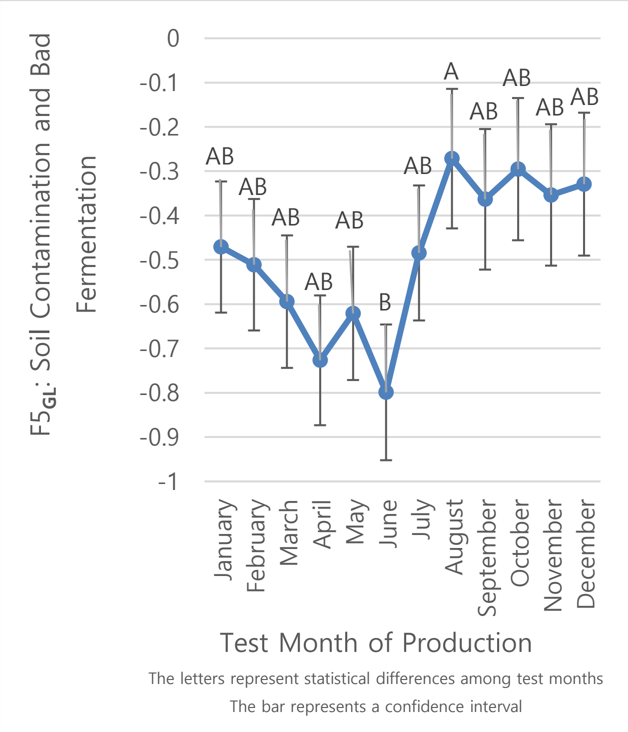


Figure 4.7: Effect of Test Month of Production for $F5_{GL}$

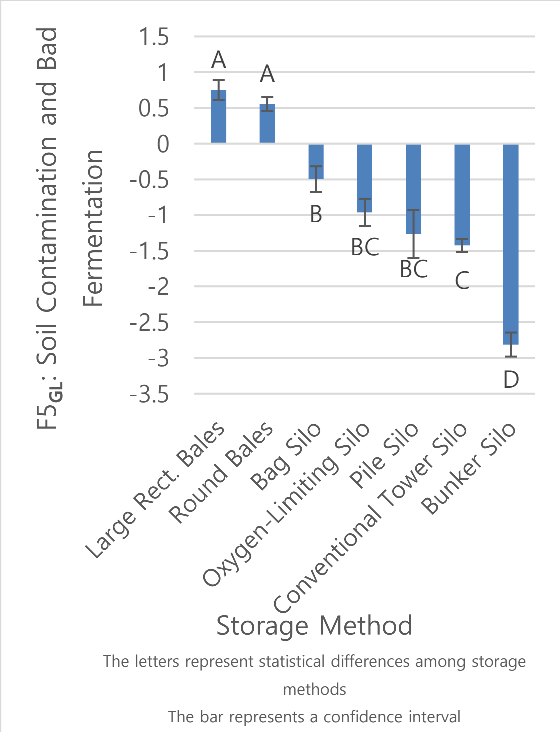


Figure 4.8: Effect of Storage Method for $F5_{GL}$

The results of the mixed model analysis for $F5_{GL}$ (**Table 4.3**) confirm the interpretation of the factor. A significative difference was observed between the test month of June (-0.80) and August (-0.27) (**Figure 4.7**). The effect of storage methods on $F5_{GL}$, showed that the round and large rectangular bales (**Figure 4.8**) had a value of 0.55 and 0.75, respectively, compared to bag silos, conventional tower silos and bunker silos, which individually showed statistical differences with -0.50, -1.43 and -2.81, respectively. Pile silos and oxygen-limiting silos bag silos have intermediated results. The statistical differences shown are likely due to the slower fermentation of bales compared to silos, which could have an impact on improving fermentation profile.

4.1.3 Mixed Effects Models to Evaluate the Impact of Grass and Legume Silage Quality on Milk Production

Using mixed effects models, the grass and legume silage quality factors and milk production and management related variables presented interesting relationships among dependent variables, which were key indicators for optimal milk production (**Table 4.3**). The threshold for p-value was set to 0.05 to signify the importance of independent variables.

4.1.3.1 Impact of Milk Production Variables and Silage Characteristics on Average Daily Milk per Cow from Grass and Legume Silage

Table 4.3 shows that increasing the proportion of corn silage, concentrate or legume from total feed increased average daily milk per cow, which is likely due to the fact that concentrate or legume from total feed are known as high sources of crude protein, which could increase energy. Similar results were observed from multiple studies for proportion of corn silage (Benchaar *et al.*, 2014), proportion of concentrate (Dewhurst *et al.*, 2003) and proportion of legume (Thomas *et al.*, 1985 and Hoffman *et al.*, 1998) from total feed. Increase in days in milk resulted in decrease milk production (**Table 4.3**). A significative difference was observed for the test month of January (32.78), October (32.57) and November (32.67), compared to June (33.92) (**Figure 4.9**).

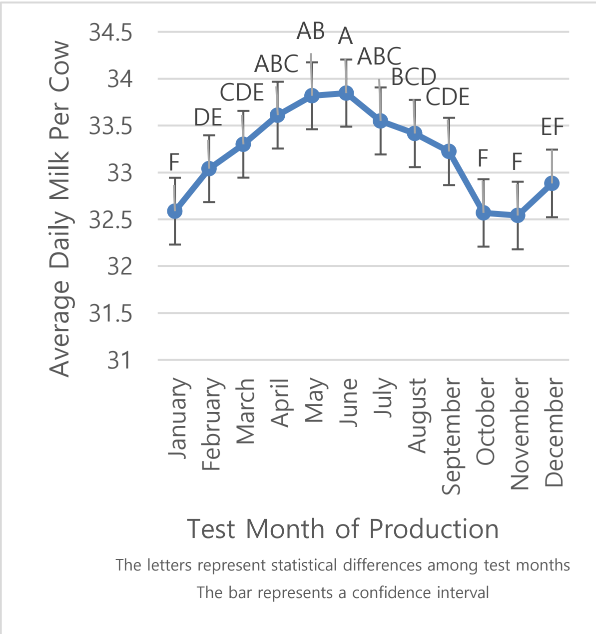


Figure 4.9: Effect of Test Month of Production for Average Daily Milk per Cow with Grass and Legume Silage

4.1.3.2 Impact of Milk Production Variables and Silage Characteristics on Average Daily Fat per Cow from Grass and Legume Silage

Similar to the increase in average daily milk per cow, an increase in the proportion of corn silage and concentrate from feed and homolactic fermentation increased fat (**Table 4.3**). This was reasonable given that an increase in milk production generally leads to an increase in fat. $F1_{GL}$ (Neutral detergent digestibility and heat damage), $F2_{GL}$ (Protein Ruminal Degradability) and $F5_{GL}$ (soil contamination and bad fermentation pattern) had no significant impact on fat. A significative difference was observed for the test month of January (1.35), compared to July (1.31) and August (1.31) (**Figure 4.10**). Similar to milk production, Average days in milk has a negative impact for fat (**Table 4.3**).

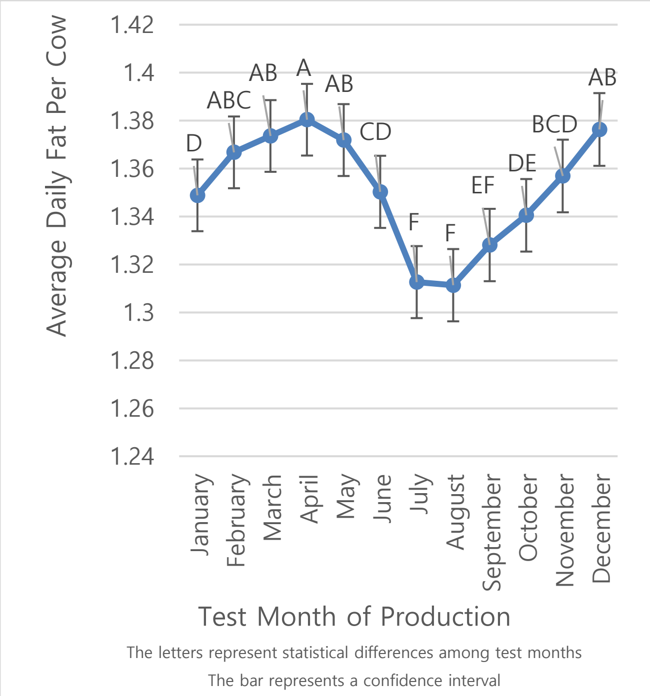


Figure 4.10: Effect of Test Month of Production for Average Daily Fat per Cow with Grass and Legume Silage

4.1.3.3 Impact of Milk production Variables and Silage Characteristics on Average Daily Protein per Cow from Grass and Legume Silage

Mixed effects model results (**Table 4.3**) showed that protein increases when proportion of corn silage, concentrate and legume silage from feed increased since these factors generally increase milk production. Benchaar *et al.* (2014) showed similar results with these factors. Unlike milk-fat production, $F4_{GL}$ (homolactic fermentation) showed no significant effect on protein. Moreover, $F1_{GL}$ (neutral detergent fiber digestibility and heat damage), $F2_{GL}$ (protein ruminal degradability) and $F5_{GL}$ (soil contamination and bad fermentation pattern) had no direct impact on protein. Average days in milk decreased average daily protein per cow (**Table 4.3**). A significative difference was observed for the test month of April (1.13) and May (1.14), compared to August (1.09) (**Figure 4.11**).

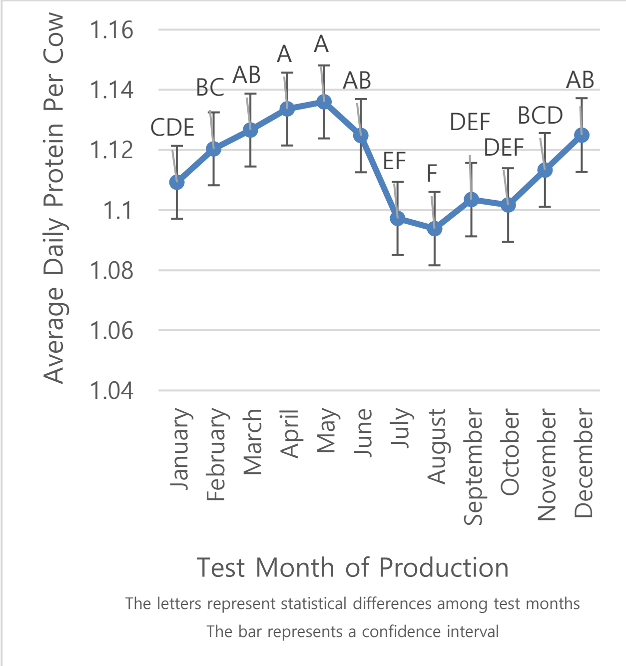


Figure 4.11: Effect of Test Month of Production for Average Daily Protein per Cow with Grass and Legume Silage

4.1.3.4 Impact of Milk production Variables and Silage Characteristics on Average Milk Urea Nitrogen from Grass and Legume Silage

All silage factors: $F1_{GL}$ (neutral detergent fiber digestibility and heat damage), $F2_{GL}$ (protein ruminal degradability), $F3_{GL}$ (legume proportion), $F4_{GL}$ (homolactic fermentation) and $F5_{GL}$ (soil contamination and bad fermentation pattern) did not have significant impact on average daily milk urea nitrogen (**Table 4.3**). Average days in milk and proportion of concentrate in feed also had no impact (**Table 4.3**). Increase in proportion of corn silage from feed showed increase in average daily milk urea nitrogen (**Table 4.3**). A significative difference was observed for the test month of January (11.01) and November (11.04), compared to May

(11.71), June (11.63) and July (11.56) (**Figure 4.12**). This is likely due to the temperature, where it is known that average milk urea nitrogen tends to increase during the summer and decrease during winter (Fatehi *et al.*, 2012).

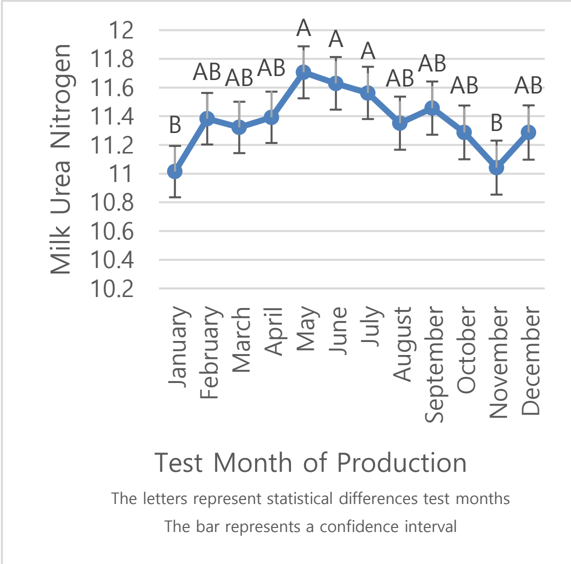


Figure 4.12: Effect of Test Month of Production for Average Milk Urea Nitrogen with Grass and Legume Silage

4.1.3.5 Impact of Milk production Variables and Silage Characteristics on Average Somatic Cell Score from Grass and Legume Silage

All silage factors: $F1_{GL}$ (neutral detergent fiber digestibility and heat damage), $F2_{GL}$ (protein ruminal degradability), $F3_{GL}$ (legume proportion), $F4_{GL}$ (homolactic fermentation) and $F5_{GL}$ (soil contamination and bad fermentation pattern) did not have a significant impact on average somatic cell score. Average days in milk and proportion of concentrate in feed also had no impact. Increase in proportion of corn silage from feed increased somatic cell score, which was reasonable considering it increases milk production and increase in milk production was known to increase somatic cell score (Zhong *et al.*, 2018). A significative difference was observed for the test month of April (3.76) and May (3.76), compared to August (4.05) (**Figure 4.13**).

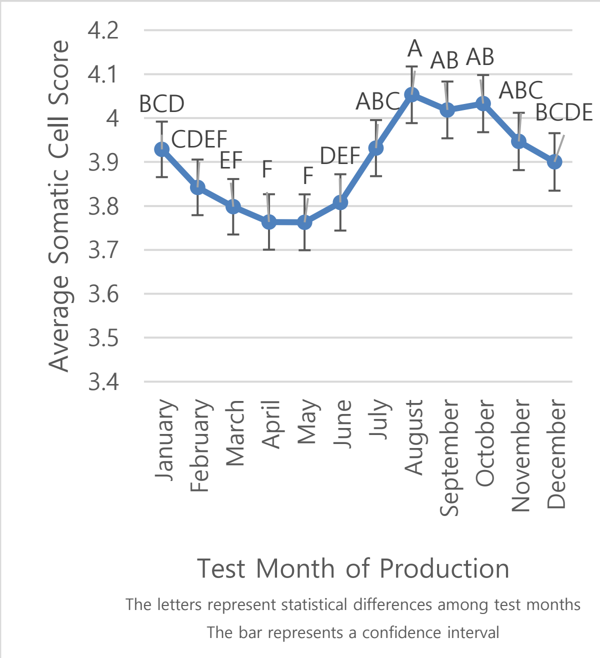


Figure 4.13: Effect of Test Month of Production for Average Somatic Cell Score with Grass and Legume Silage

4.2 Interpretation of Corn Silage Quality Imputation, Factors and Relationship with Milk Production Management

4.2.1 Silage Quality Imputation Results

Within the milk production dataset, 7,782 corn silage samples had missing input parameters and therefore did not impute the rest of the silage variables. As a result, final datasets include 382 herds and 4,116 samples for the milk production dataset with mainly corn silage (**Figure 4.1**). The accuracy metrics for the models were not high, especially the R-square for 1, 2 - Propanediol for corn silage dataset. Various feature selection techniques were explored, but they did not improve the R-Square. Various regression algorithms were explored to identify the algorithm with ideal prediction accuracy of corn silage qualities. Despite having some algorithms perform better for silage particular variables with particular metrics, an algorithm that generally performs well for all silage variables was chosen for efficiency. Highest accuracy for data imputation was found using ensemble algorithms from supervised machine-learning had a performance of approximately 99 percent average for corn silage, respectively (**Table 4.4**). More specifically, regressor chain based AdaBoost were the best algorithm for corn silage. The imputed silage dataset projected expected silage characteristics similar to the training dataset.

Table 4.4: Machine Learning Model Prediction Metrics for Corn Silage Dataset

Variable	Model	Metric				
		Mean	**MAE[1]**	**MAE / Mean**	**RMSE[2]**	**R^2**
Rumen Digestible Neutral Detergent Fiber at 30 Hour in Vitro	Linear Regression	0.88	0.02	0.02	0.03	0.67
	Robust (Huber) Regression	0.88	0.01	0.02	0.03	0.51
	PLS (Partial Least Square) Regression	59.70	0.02	0.02	0.03	0.62
	ElasticNet Regression	0.88	0.02	0.02	0.03	0.66
	KNN (K Nearest Neighbour)	0.88	0.02	0.02	0.03	0.56
	SVR (Support Vector Regression)	59.70	1.86	0.03	3.45	0.53
	Meta (Bagging) Estimator	59.70	0.65	0.01	0.92	0.96
	Random Forest	59.70	1.27	0.02	1.74	0.88
	AdaBoost	**0.88**	**0.01**	**0.0002**	**0.06**	**0.99**
Rumen Digestible Neutral Detergent Fiber at 120 Hour in Vitro	Linear Regression	1.00	0.03	0.03	0.04	0.65
	Robust (Huber) Regression	1.00	0.02	0.02	0.04	0.53
	PLS (Partial Least Square) Regression	71.67	0.03	0.03	0.04	0.60
	ElasticNet Regression	1.00	0.03	0.03	0.04	0.64
	KNN (K Nearest Neighbour)	1.00	0.03	0.03	0.04	0.55
	SVR (Support Vector Regression)	71.67	2.38	0.03	4.32	0.53
	Meta (Bagging) Estimator	71.67	0.83	0.01	1.17	0.96
	Random Forest	71.67	1.64	0.02	2.25	0.87
	AdaBoost	**1.00**	**0.01**	**0.0002**	**0.08**	**0.99**
Rumen Digestible Neutral Detergent Fiber at 240 Hour in Vitro	Linear Regression	1.03	0.03	0.03	0.04	0.65
	Robust (Huber) Regression	1.03	0.02	0.02	0.05	0.54
	PLS (Partial Least Square) Regression	74.09	0.03	0.03	0.04	0.60
	ElasticNet Regression	1.03	0.03	0.03	0.04	0.64
	KNN (K Nearest Neighbour)	1.03	0.03	0.03	0.05	0.55
	SVR (Support Vector Regression)	74.09	2.44	0.03	4.48	0.52
	Meta (Bagging) Estimator	74.09	0.85	0.01	1.19	0.96
	Random Forest	74.09	1.68	0.02	2.30	0.87
	AdaBoost	1.03	0.01	**0.0002**	0.10	**0.99**
pH	Linear Regression	3.92	0.07	0.01	0.10	0.56
	Robust (Huber) Regression	3.92	0.07	0.01	0.09	0.56
	PLS (Partial Least Square) Regression	3.92	0.08	0.02	0.12	0.46
	ElasticNet Regression	3.92	0.07	0.01	0.11	0.55
	KNN (K Nearest Neighbour)	3.92	0.07	0.02	0.11	0.53
	SVR (Support Vector Regression)	3.92	0.08	0.02	0.11	0.49
	Meta (Bagging) Estimator	3.92	0.07	0.01	0.10	0.64
	Random Forest	3.92	0.06	0.01	0.09	0.65
	AdaBoost	3.92	0.0001	**0.00003**	0.003	**0.99**
Lactic Acid	Linear Regression	0.18	0.02	0.12	0.03	0.69
	Robust (Huber) Regression	0.18	0.02	0.11	0.02	0.61
	PLS (Partial Least Square) Regression	3.57	0.02	0.15	0.03	0.61
	ElasticNet Regression	0.18	0.02	0.13	0.03	0.68
	KNN (K Nearest Neighbour)	0.18	0.02	0.12	0.03	0.70
	SVR (Support Vector Regression)	3.57	0.78	0.22	1.01	0.62
	Meta (Bagging) Estimator	3.57	0.29	0.08	0.38	0.94
	Random Forest	3.57	0.61	0.17	0.82	0.75
	AdaBoost	0.18	0.003	**0.0009**	0.02	**0.99**
Acetic Acid	Linear Regression	0.14	0.02	0.15	0.02	0.49
	Robust (Huber) Regression	0.14	0.02	0.14	0.02	0.49
	PLS (Partial Least Square) Regression	2.22	0.02	0.15	0.02	0.45
	ElasticNet Regression	0.14	0.02	0.15	0.02	0.47
	KNN (K Nearest Neighbour)	0.14	0.02	0.16	0.02	0.44
	SVR (Support Vector Regression)	2.22	0.61	0.27	0.81	0.48
	Meta (Bagging) Estimator	2.22	0.62	0.27	0.81	0.48
	Random Forest	2.22	0.46	0.20	0.60	0.71
	AdaBoost	0.14	0.0006	**0.0003**	0.01	**0.99**
Propionic Acid	Linear Regression	0.05	0.009	0.18	0.01	0.48
	Robust (Huber) Regression	0.05	0.008	0.16	0.01	0.47
	PLS (Partial Least Square) Regression	0.27	0.01	0.20	0.01	0.39
	ElasticNet Regression	0.05	0.01	0.19	0.01	0.46
	KNN (K Nearest Neighbour)	0.05	0.01	0.21	0.01	0.34
	SVR (Support Vector Regression)	0.27	0.08	0.32	0.11	0.48
	Meta (Bagging) Estimator	0.27	0.08	0.31	0.11	0.51
	Random Forest	0.27	0.06	0.24	0.08	0.70

Variable	Model	Metric				
		Mean	MAE[1]	MAE / Mean	RMSE[2]	R^2
	AdaBoost	0.05	0.00001	**0.0002**	0.001	**0.99**
Propanediol - 1, 2	Linear Regression	0.13	0.01	0.13	0.03	0.27
	Robust (Huber) Regression	0.13	0.01	0.12	0.03	0.12
	PLS (Partial Least Square) Regression	1.81	0.02	0.15	0.03	0.16
	ElasticNet Regression	0.13	0.01	0.14	0.03	0.23
	KNN (K Nearest Neighbour)	0.13	0.01	0.14	0.02	0.24
	SVR (Support Vector Regression)	1.81	0.45	0.25	0.67	0.20
	Meta (Bagging) Estimator	1.81	0.45	0.24	0.67	0.21
	Random Forest	1.81	0.41	0.23	0.57	0.41
	AdaBoost	0.13	0.0003	**0.0002**	0.003	**0.99**
Starch Digestibility	Linear Regression	1.00	0.02	0.02	0.03	0.78
	Robust (Huber) Regression	1.00	0.02	0.02	0.03	0.74
	PLS (Partial Least Square) Regression	70.12	0.03	0.03	0.04	0.69
	ElasticNet Regression	1.00	0.03	0.03	0.03	0.76
	KNN (K Nearest Neighbour)	1.00	0.04	0.04	0.05	0.60
	SVR (Support Vector Regression)	70.12	2.66	0.03	3.62	0.76
	Meta (Bagging) Estimator	70.12	2.64	0.03	3.61	0.77
	Random Forest	70.12	2.67	0.03	3.46	0.78
	AdaBoost	1.00	0.009	**0.0001**	0.17	**0.99**

[1]MAE: Mean Absolute Error
[2]RMSE: Root Mean Squared Error

4.2.2 Silage Quality Factor Analysis

Corn silage variables were grouped into five factors:

- $F1_C$: Starch Concentration and Maturity
- $F2_C$: Homolactic Fermentation, Starch Digestibility and Fermentation Length
- $F3_C$: Neutral Detergent Fiber Digestibility
- $F4_C$: Protein Rumen Degradability
- $F5_C$: Heterolactic and Other Secondary Fermentations

Similar to the steps taken for grass and legume silage, initial step was to analyze the impact of test date, region and storage method on factors derived from the factor loadings (**Table 4.5**). JMP statistical software was used to perform statistical analysis. For each factor, Tukey results with least square means were generated to observe statistical differences among the variables.

Table 4.5: Factor Loadings for Corn Silage

Factors	$F1_C$	$F2_C$	$F3_C$	$F4_C$	$F5_C$	Communalities
Dry Matter (%)	40.21	-20.98	-2.66	-7.84	-12.74	0.23
Crude Protein (DM %)	32.99	-32.22	16.06	87.86	4.23	1.01
Soluble Protein (DM %)	-1.59	71.38	-12.76	49.95	38.25	0.92
Soluble Protein (CP %)	17.67	86.65	-19.90	-3.70	33.71	0.94
Acid Detergent Insoluble Crude Protein (DM %)	-64.14	-59.14	7.01	8.75	-5.75	0.78

Factors	F1c	F2c	F3c	F4c	F5c	Communalities
Acid Detergent Insoluble Crude Protein (CP %)	-37.89	-32.70	-7.08	-68.31	-11.78	0.74
Neutral Detergent Insoluble Crude Protein (DM %)	-59.86	-60.67	23.73	20.12	-16.35	0.85
Neutral Detergent Insoluble Crude Protein (CP %)	-51.72	-53.21	17.42	-38.41	-24.53	0.79
Rumen Degradable Protein (DM %)	-24.56	17.03	4.44	90.82	24.66	0.98
Rumen Degradable Protein (CP %)	17.65	86.58	-19.83	-3.64	33.68	0.93
Acid Detergent Fiber (DM %)	-97.16	-3.33	-13.68	-7.57	3.35	0.97
Acid Detergent Fiber (NDF %)	-54.83	7.50	-45.31	-16.07	1.72	0.54
Neutral Detergent Fiber (DM %)	-95.45	-7.27	-0.03	-2.60	3.28	0.92
Lignin (DM %)	-73.44	-35.52	-39.71	3.98	10.84	0.84
Lignin (NDF %)	-18.05	-43.36	-55.18	6.79	10.85	0.54
Crude Fat (DM %)	-2.61	-0.98	14.69	25.33	55.26	0.39
Non-fibrous Carbohydrate (DM %)	94.87	7.45	5.65	-15.82	-6.51	0.94
Ash (DM %)	-8.37	9.42	-43.59	0.03	-12.26	0.22
Starch (DM %)	94.93	-2.56	-5.84	-15.68	-1.14	0.93
Rumen Digestible Neutral Detergent Fiber at 30 Hour in Vitro (NDF %)	-1.14	-7.83	97.73	6.38	1.60	0.97
Rumen Digestible Neutral Detergent Fiber at 120 Hour in Vitro (NDF %)	-3.74	-8.14	97.69	6.48	1.85	0.97
Rumen Digestible Neutral Detergent Fiber at 240 Hour in Vitro (NDF %)	-2.18	-7.36	97.79	6.66	1.63	0.97
pH	0.07	-76.03	-5.92	5.25	10.27	0.59
Starch Digestibility (DM %)	-12.15	68.85	3.04	8.69	-1.19	0.50
Lactic Acid (DM %)	-5.10	88.45	-0.45	16.30	-24.63	0.87
Acetic Acid (DM %)	-22.46	18.13	16.84	5.36	88.62	0.90
Propionic Acid (DM %)	-11.53	6.53	3.08	-3.04	68.08	0.48
1, 2 – Propanediol (DM %)	7.88	-2.35	-7.85	7.37	41.75	0.19
Eigenvalues	7.78	5.23	4.31	2.60	1.97	N / A
Explained Variance	22.14	19.74	14.39	9.74	8.58	N / A

$F1_C$: Starch Concentration and Plant Maturity
$F2_C$: Homolactic Fermentation, Starch Digestibility and Fermentation Length
$F3_C$: Neutral Detergent Fiber Digestibility
$F4_C$: Protein Rumen Degradability
$F5_C$: Heterolactic and Other Secondary Fermentations
N / A: Not Applicable
Green and red colors represent positive and negative values, respectively.

Table 4.6: Corn Silage Variables Used in Mixed Effects Modelling and Their Effects

Independent Variables	Dependent Variables	Estimates (+, -)	p-values
Average Daily Milk per Cow (kg)	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	**-0.05**	< 0.05
	Proportion of Concentrate from Feed	**1.20**	< 0.05
	Total Proportion of Grass and Legume Silage	**-1.31**	< 0.05

Independent Variables	Dependent Variables	Estimates (+, -)	p-values
	$F1_C$	**0.04**	0.15
	$F2_C$	-0.02	0.59
	$F3_C$	0.001	0.96
	$F4_C$	**-0.23**	< 0.05
	$F5_C$	**0.37**	< 0.05
Average Daily Fat per Cow (kg)	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	**-0.001**	< 0.05
	Proportion of Concentrate from Feed	**0.05**	< 0.05
	Total Proportion of Grass and Legume Silage	-0.013	0.58
	$F1_C$	0.0004	0.73
	$F2_C$	-0.0001	0.92
	$F3_C$	0.001	0.37
	$F4_C$	-0.004	0.17
	$F5_C$	**0.01**	< 0.05
Average Daily Protein per Cow (kg)	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	**-0.001**	< 0.05
	Proportion of Concentrate from Feed	0.0005	0.98
	Total Proportion of Grass and Legume Silage	**-0.06**	< 0.05
	$F1_C$	0.001	0.12
	$F2_C$	0.0005	0.65
	$F3_C$	-0.0004	0.72
	$F4_C$	**-0.006**	< 0.05
	$F5_C$	**0.007**	< 0.05
Average Somatic Cell Score	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	0.19
	Average Days in Milk	**0.001**	< 0.05
	Proportion of Concentrate from Feed	-0.34	0.06
	Total Proportion of Grass and Legume Silage	**-0.51**	< 0.05
	$F1_C$	0.005	0.58
	$F2_C$	0.003	0.75
	$F3_C$	0.005	0.61
	$F4_C$	0.01	0.51
	$F5_C$	-0.02	0.25
Average Daily Milk Urea Nitrogen	Herd	N/A	< 0.05
	Test Year of Production	N/A	< 0.05
	Test Month of Production	N/A	< 0.05
	Region	N/A	< 0.05
	Average Days in Milk	-0.001	0.42
	Proportion of Concentrate from Feed	**-1.38**	< 0.05
	Total Proportion of Grass and Legume Silage	**-2.34**	< 0.05
	$F1_C$	-0.03	0.26
	$F2_C$	0.060	0.08
	$F3_C$	0.004	0.91
	$F4_C$	-0.11	0.09
	$F5_C$	0.01	0.80

$F1_C$: Starch Concentration and Maturity
$F2_C$: Homolactic Fermentation, Starch Digestibility and Fermentation Length
$F3_C$: Neutral Detergent Fiber Digestibility
$F4_C$: Protein Rumen Degradability
$F5_C$: Heterolactic and Other Secondary Fermentations
N/A: Not Available.
Estimates are highlighted in green for positive values and red for negative values.

4.2.2.1 Impact of Silage Management Variables on Starch Concentration and Maturity with Corn Silage

The factor $F1_C$ (**Table 4.5**) was defined as an index related to starch concentration because the factor loading (**Table 4.5**) for starch was 94.93. The factor loadings for lignin (%DM), ADF (%DM and %NDF), CP bound to ADF (%DM), NDF (DM %) and CP bound to NDF (%DM and %CP) were the opposite of starch concentration and plant maturity and significantly negative. DM, ADF and CP bound to ADF (Bal *et al.*, 1997) was recognized as an indicator of plant maturity.

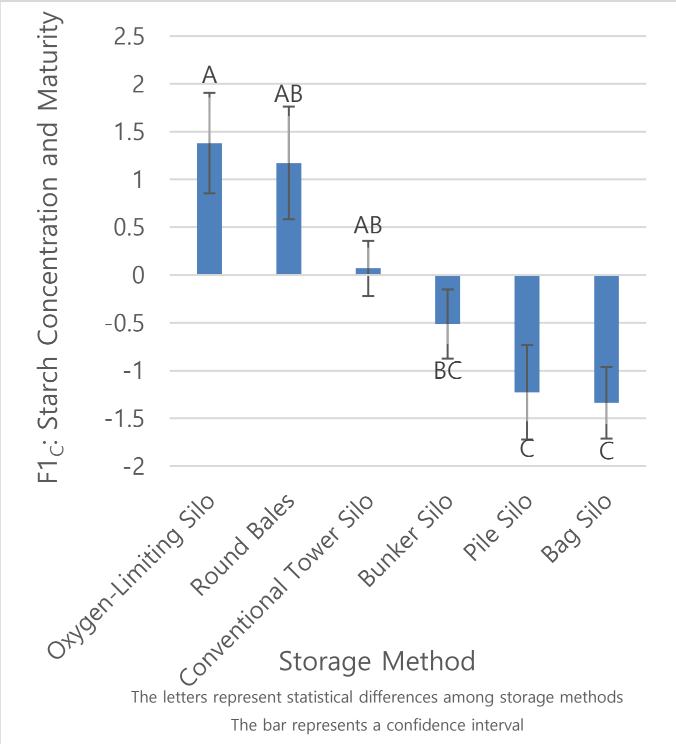


Figure 4.14: Effect of Storage Method for $F1_C$

The results of the mixed model analysis for $F1_C$ (**Table 4.6**) confirm the interpretation of the factor. The effect of storage methods on $F1_C$, showed that the oxygen-limiting silos (**Figure 4.14**) had a value of 1.38, compared to pile silos and bag silos with -1.23 and -1.34, respectively. Round bales, bunker silos and conventional tower silos have intermediated results. The high plant maturity for oxygen-limiting silos is likely since it requires higher DM of silage.

4.2.2.2 Impact of Silage Management Variables on Homolactic Fermentation, Starch Digestibility and Fermentation Length with Corn Silage

The factor F2c (**Table 4.5**) was defined as an index related to homolactic fermentation and starch digestibility because the factor loadings (**Table 4.5**) for lactic acid and starch digestibility were 88.45 and 68.85, respectively. The factor loadings for lignin (%DM), CP bound to ADF (%DM) and CP bound to NDF (%DM and %CP) were the opposite of F2c and significantly negative. pH (Opinya, 2019) was recognized as an indicator of optimal acid production, which is related to fermentation length. In addition, soluble protein (Windle *et al.,* 2014) was recognized as an indicator for fermentation length.

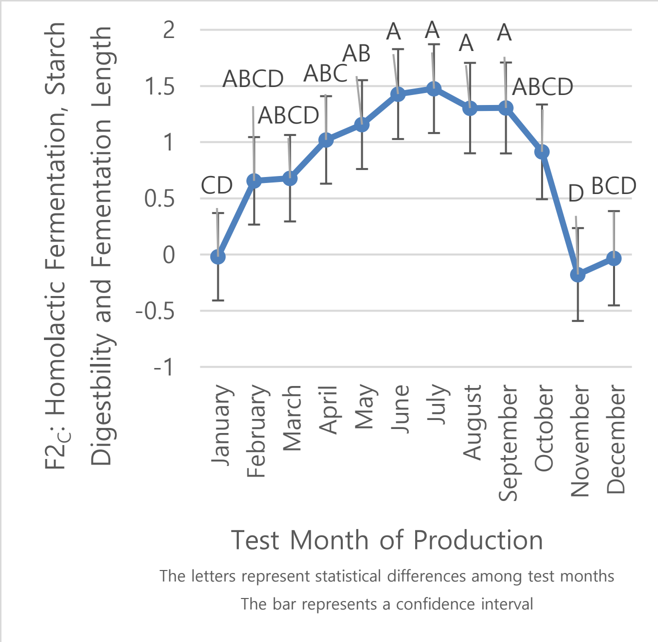


Figure 4.15: Effect of Test Month of Production for F2c

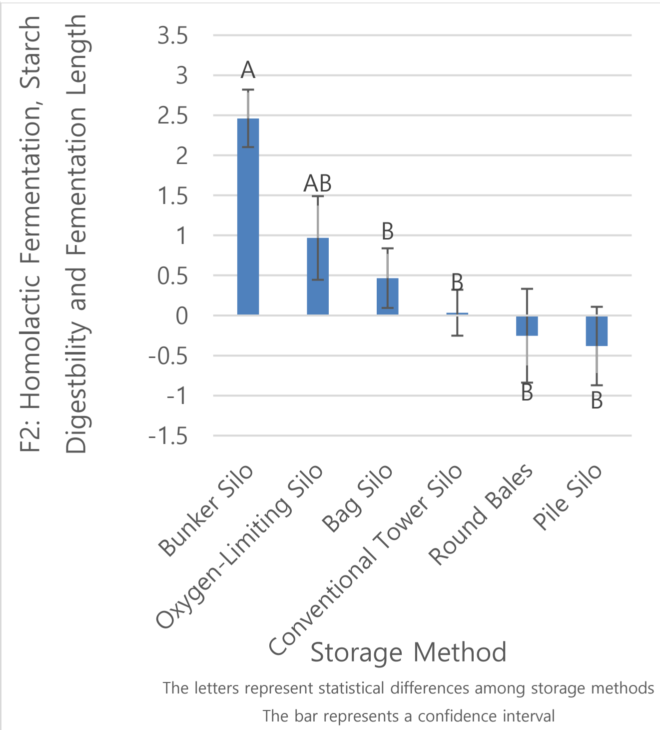


Figure 4.16: Effect of Storage Method of Production for F2c

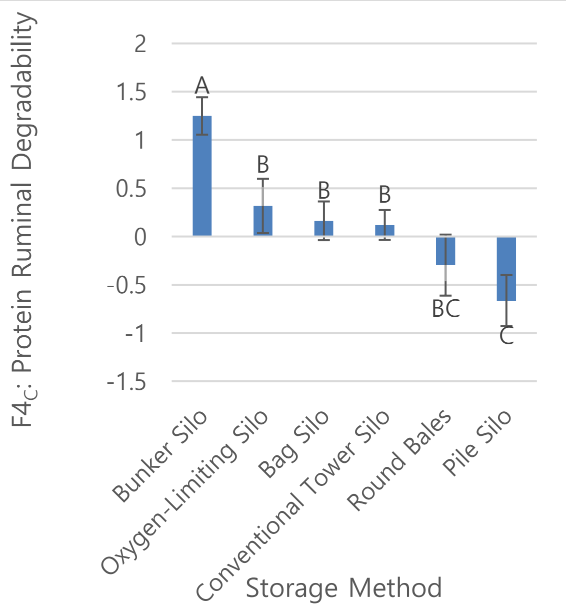
The results of the mixed model analysis for F2c (**Table 4.6**) confirm the interpretation of the factor. A significative difference was observed in June (1.43), July (1.48), August (1.30) and September (1.30) compared to and November (-0.18) (**Figure 4.15**). Also, the effect of storage methods on F2c, showed that bunker silos (**Figure 4.16**) had a value of 2.46, compared to other storage methods such as bag silos, conventional tower silos, pile silos and round bales with 0.47, 0.35, -0.38 and -0.25, respectively. Oxygen-limiting silos have intermediated results. The higher values for bunker silos are likely due to the longer fermentation time (Oney et al., 2018). In addition, the lower value of round bales is likely due to having slower fermentation caused by being loosely packaged and having more oxygen (Schick, 2019).

4.2.2.3 Impact of Silage Management Variables on Neutral Detergent Fiber Digestibility with Corn Silage

The factor $F3_C$ (**Table 4.5**) was defined as an index related to neutral detergent fiber digestibility (NDFD) because the factor loadings (**Table 4.5**) for rumen digestible neutral detergent fiber digestibility (NDFD) in vitro (30, 120, 240 hours) were 97.73, 97.69 and 97.79 respectively. The factor loadings for lignin (%NDF), ADF (%NDF) and ash (%DM) were the opposite of NDFD and significantly negative.

4.2.2.4 Impact of Silage Management Variables on Protein Ruminal Degradability with Corn Silage

The factor $F4_C$ (**Table 4.5**) was defined as an index related to protein ruminal degradability because the factor loadings (**Table 4.5**) for rumen degradable protein (%DM), crude protein (%DM) and soluble protein (%DM) were 90.82, 87.86 and 49.95, respectively. The factor loading for CP bound to ADF (%CP) was the opposite of rumen degradable protein, crude protein and soluble protein and significantly negative.



The letters represent statistical differences among storage methods
The bar represents a confidence interval

Figure 4.17: Effect of Storage Method for $F4_C$

The results of the mixed model analysis for $F4_C$ (**Table 4.5**) confirm the interpretation of the factor. The effect of storage methods on $F4_C$, showed that bunker silos (**Figure 4.17**) had a value of 1.25, compared to other storage methods such as oxygen-limiting silos, conventional tower silos and bag silos with 0.32, 0.12 and 0.16, respectively. Pile silos had a value of -0.66, which was statistically different from other storage methods. Round bales have intermediated results.

4.2.2.5 Impact of Silage Management Variables on Heterolactic and Other Secondary Fermentations with Corn Silage

The factor $F5_C$ (**Table 4.5**) was defined as an index related to heterolactic and other secondary fermentations because the factor loadings (**Table 4.5**) for acetic acid (%DM), propionic acid (%DM), crude fat and propanediol 1, 2 (%DM) were 88.62, 68.08, 55.26 and 41.75, respectively.

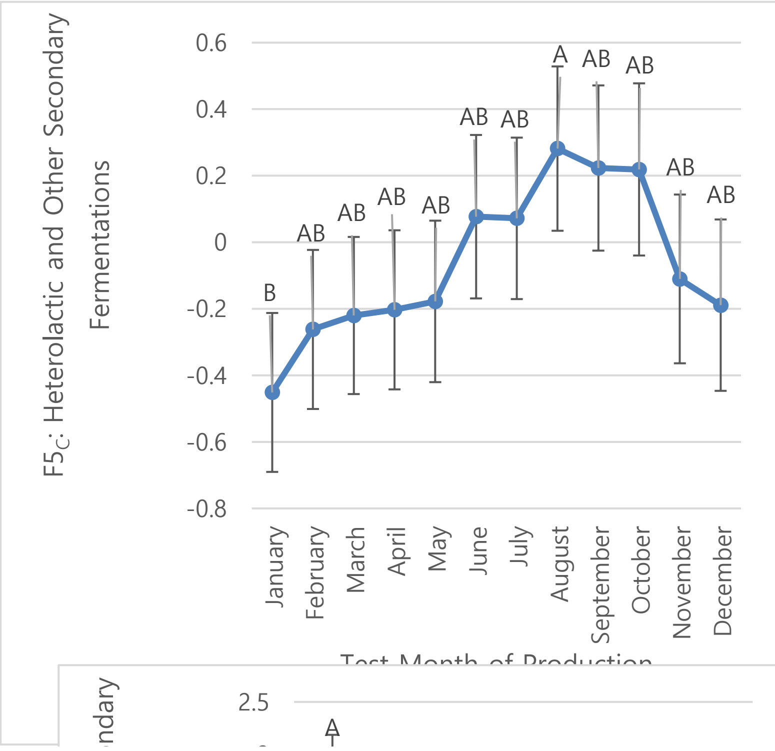
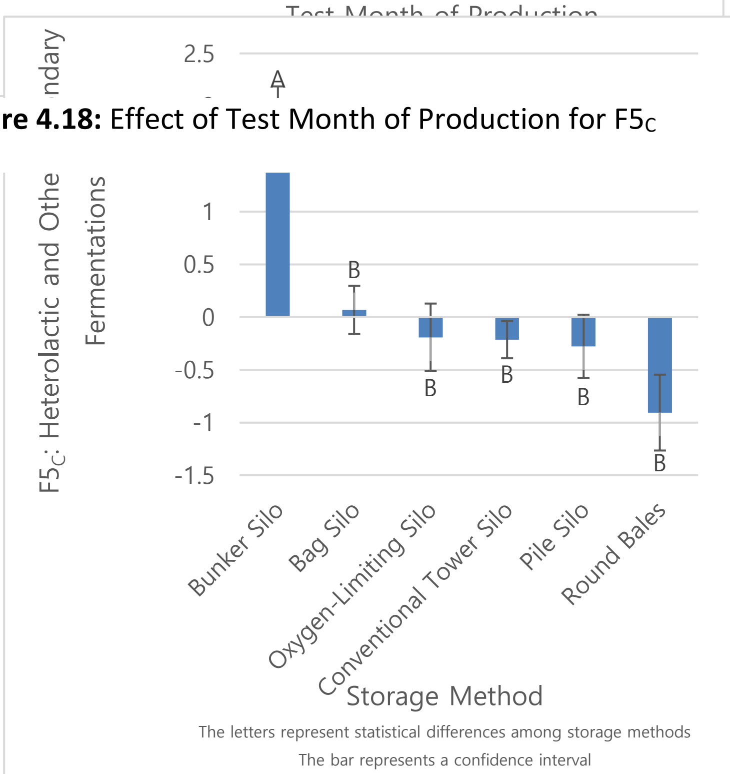


Figure 4.18: Effect of Test Month of Production for $F5_C$

The results of the mixed model analysis for $F5_C$ (**Table 4.5**) confirm the interpretation of the factor. A significative difference was observed in January (-0.45) compared to August (0.28) (**Figure 4.18**). The effect of storage methods on $F5_C$, showed that bunker silos (**Figure 4.19**) had a value of 1.97, compared to other storage methods such as bag silos, conventional tower silos, pile silos,



oxygen-limiting silos and round bales with 0.07, -0.21, -0.28, -0.19 and -0.91, respectively.

Figure 4.19: Effect of Storage Method for $F5_C$

4.2.3 Mixed Effects Models to Evaluate the Impact of Corn Silage Quality on Milk Production

Using mixed effects models, the corn silage quality factors and milk production and management related variables presented significant relationships among dependent variables, which were key indicators for optimal milk production (**Table 4.6**). The threshold for p-value was set to 0.05 to signify the importance of independent variables.

4.2.3.1 Impact of Milk production Variables and Silage Characteristics on Average Daily Milk per Cow from Corn Silage

Mixed-effects model results indicated that the increase in $F5_C$ (heterolactic and other secondary fermentations) increase milk production (**Table 4.6**). There is a possibility that this is due to the creation of lactic, acetic and propionic acid that increases aerobic stability and other study has shown similar results (Oliveira *et al*., 2016). In addition, **Table 4.6** demonstrated that increase in proportion of concentrate from feed increases milk production. Similar behavior was observed from a study conducted by Dewhurst *et al*., (2003). However, an increase in $F4_C$ (protein ruminal degradability) and grass and proportion of legume silage from feed decreases milk production. Average days in milk had a negative impact on milk production. Proportion of concentrate from feed and $F2_C$ (homolactic fermentation, starch digestibility and fermentation length) have no direct impact on milk production from our results (**Table 4.6**). A significative difference was observed for the test month of October (32.57) and November (32.67), compared to May (34.00) and June (33.92) (**Figure 4.20**). The difference is due to the higher risk of heat damage during summer months compared to the winter month like observed by Bernardes *et al.* (2018).

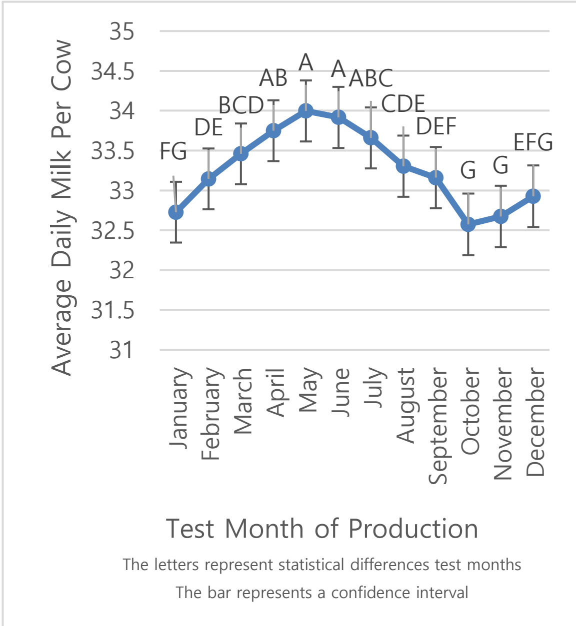


Figure 4.20: Effect of Test Month of Production for Average Daily Milk per Cow with Corn Silage

4.2.3.2 Impact of Milk production Variables and Silage Characteristics on Average Daily Fat per Cow from Corn Silage

Table 4.6 demonstrated that increase in the proportion of concentrate, $F5_C$ (heterolactic and other secondary fermentations) increased milk-fat. There is a possibility that this is due to creation of lactic, acetic and propionic acids that promote aerobic stability. Similar results were observed by a study from Oliveira *et al*., (2016). Average days in milk negatively impacts fat. $F1_C$ (starch concentration and plant maturity), $F2_C$ (homolactic fermentation, fermentation length and starch digestibility) proportion of grass legume silage from feed, length, $F3_C$ (neutral detergent digestibility) and $F4_C$ (protein ruminal degradability) has little to no impact according to the results. A significative difference was observed for the test month of April (1.38) and December (1.38), compared and July (1.31) and August (1.31) (**Figure 4.21**). In addition, June (1.35) was statistically difference from the rest of the test months.

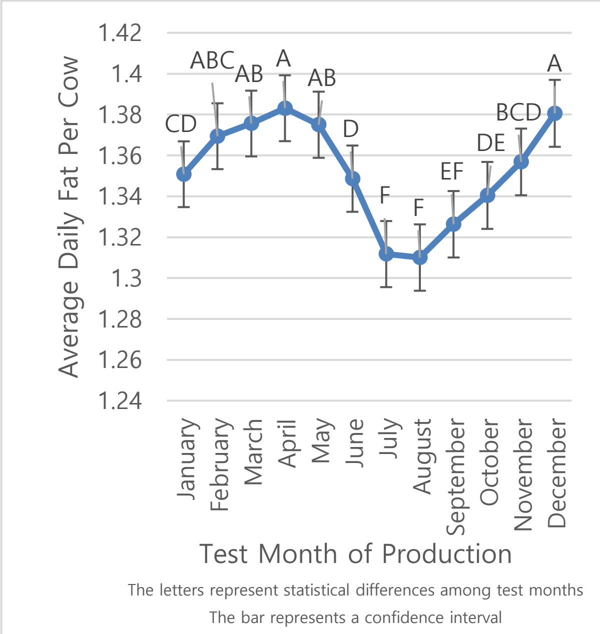


Figure 4.21: Effect of Test Month of Production for Average Daily Fat per Cow with Corn Silage

4.2.3.3 Impact of Milk production Variables and Silage Characteristics on Average Daily Protein per Cow from Corn Silage

The results indicated that an increase in $F5_C$ (heterolactic and other secondary fermentations) increased protein (**Table 4.6**). In addition, increase in average days in milk and protein ruminal degradability was shown to decrease protein (**Table 4.6**). A study from Erdman *et al*. (1983) also showed that decrease in $F4_C$ (protein ruminal degradability) decreased protein. In addition, an increase in the proportion of grass and legume silage in feed decreased protein as expected. Proportion of concentrate from feed, $F1_C$ (starch concentration and plant maturity), $F2_C$ (homolactic fermentation, fermentation length and starch digestibility) and $F3_C$

(neutral detergent digestibility) had no impact on protein. A significative difference was observed for the test month of April (1.14) and May (1.14), compared to August (1.09) (**Figure 4.22**).

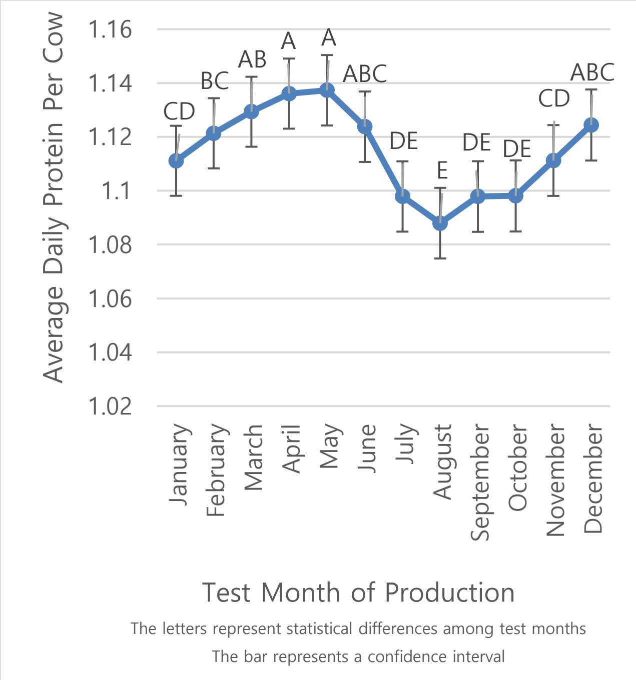


Figure 4.22: Effect for Test Month of Production and Average Daily Protein per Cow with Corn Silage

4.2.3.4 Impact of Milk production Variables and Silage Characteristics on Average Milk Urea Nitrogen from Corn Silage

Mixed-effects results indicated that increase in proportion of concentrate or grass and legume silage decreased average daily milk urea nitrogen (**Table 4.6**). Average days in milk and all corn silage factors did not have significant impact on average daily milk urea nitrogen. A significative difference was observed for the test month of January (10.76), compared to May (11.58) (**Figure 4.23**). The difference is due to the higher risk

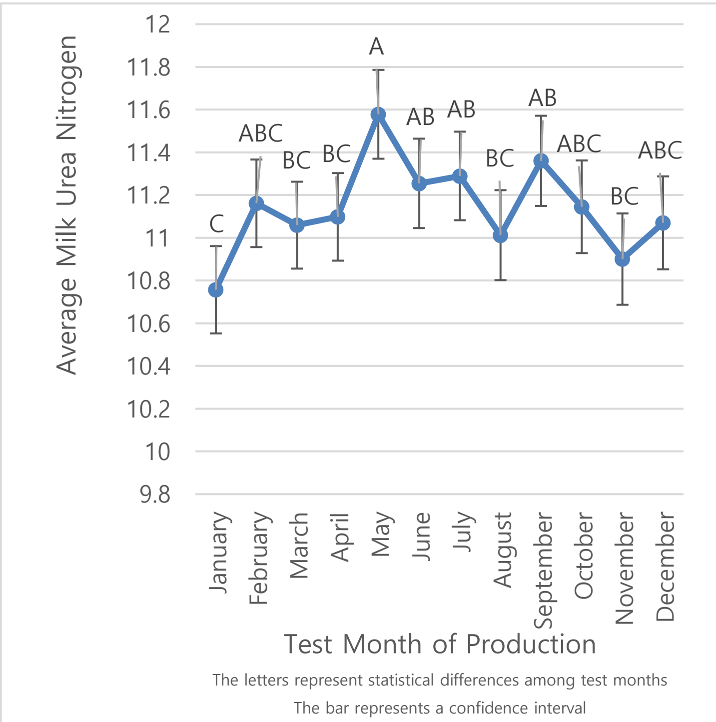


Figure 4.23: Effect of Test Month of Production for Average Milk Urea Nitrogen Corn Silage

of heat damage during summer months compared to the winter month like observed by Bernardes *et al.* (2018).

4.2.3.5 Impact of Milk production Variables and Silage Characteristics on Average Somatic Cell Score from Corn Silage

Mixed-effects results show that increase in proportion of grass and legume silage decreases average somatic cell score (**Table 4.6**). Similar results were shown from a study that observed a decrease in crude protein led to increase in somatic cell count (Litwinczuk *et al.*, 2011). More specifically, legume is known to increase crude protein, meaning increase in legume would decrease somatic cell count, which in our study is interpreted using the score.

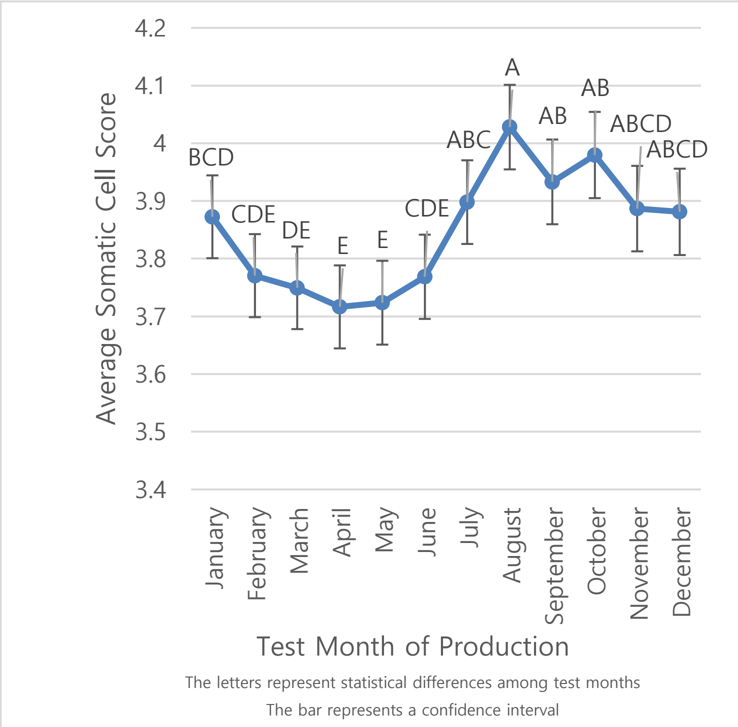


Figure 4.24: Effect of Test Month of Production for Average Somatic Cell Score with Corn Silage

Increase in average days in milk increase average somatic cell score (**Table 4.6**), which was also observed from Yoon *et al*., 2004. Proportion of concentrate from feed and all corn silage factors did not have significant impact on average somatic cell score (**Table 4.6**). A significative difference was observed for the test month of April (3.72) and May (3.72), compared to August (4.03) (**Figure 4.24**). This is likely due to higher risk of bacterial infection during summer, which increases somatic cell counts (Hammami *et al.*, 2013).

5. Conclusion

The principal objective of this study was to introduce the relationship between silage characteristics and milk production. It is well known that silage quality has an impact on milk production, but there is a lack of comprehensive approaches to describe this impact. More specifically, individual studies have shown significant impacts of key silage quality factors (such as digestibility, fermentation characteristics and nutritional constituents) on milk yield and its composition, but the nature of those relationships requires domain knowledge, and is not easily disseminated to producers in a comprehensible fashion. In addition, simultaneous data analysis of digestibility, fermentation characteristics, nutritional constituents and their dependencies on milk, fat, and protein yield is necessary. In this research, a data-driven approach was proposed: a rigorous data analysis through the combination of machine learning and statistical models.

It is important to note that raw data are not to be used directly for data analysis. Data collected from different fields and herds need to undergo extensive data cleaning processes to ensure their validity. Multi-layer outlier detection is suggested to analyze outliers for both individual variables, and each sample that consists of all variables, in order to filter values that are numerically out of range, or statistically or biologically questionable.

Datasets that combine field and herd data are not easily available. Therefore, one approach to work with such datasets is to artificially generate field data within herd datasets. Machine learning is the preferred method if high numerical precision and accuracy is the priority, compared to statistical inference, which focuses more on understanding the relationships among variables. For the silage quality variable predictions, multi-input and multi-output (MIMO) regression was performed since generally, it is computationally less expensive, executes much faster, and also considers the whole sample as inputs and outputs (as opposed to individual variables). These reasons are important since silage consists of a combination of key variables, and low computational cost is necessary for potential extension of this research with more data and deployment in the industry. Ensemble methods such as meta-estimator with Extra Tree algorithm as base regressor and regressor chain based AdaBoost algorithm with Extra Tree algorithm as the base regressor were used for grass and legume silage and corn silage, respectively. However, deep learning algorithms could be considered if a dataset is much larger and more complex (multi-dimensional variables, images, audio, etc.).

Clustering silage variables to define key silage factors would help the animal feed industry since, not only is the analysis of individual variable time consuming, but the results are frequently only interpretable by domain experts. Extensive research and development on the proposed approach would facilitate producers in understanding the quality of their silage and how it can impact milk production in their herds.

Finally, if the results of this preliminary research are deemed useful by the industry, the option exists to extend and promote them through the development of a decision-support tool for silage quality evaluation, thereby providing a practical guide to industry experts and producers for improved dairy profitability.

6. Reference

Agabriel, C., A. Cornu, C., Sibra, C., Grolier, P. and B. Martin. 2007. Tanker Milk Variability According to Farm Feeding Practices: Vitamins a and E, Carotenoids, Color, and Terpenoids. *J. Dairy Sci.* 90: 4884 - 96.

Aghamohammadi, M., Haine, D., Kelton, D. F., Barkema, H. W., Hogeveen, H. G., Keefe, P. and Dufour, S. 2018. Herd - Level Mastitis - Associated Costs on Canadian Dairy Farms. *Front. Vet. Sci.* 5: 100.

Al-Mabruk, R. M., Beck, N. F. and Dewhurst, R. J. 2004. Effects of Silage Species and Supplemental Vitamin E on the Oxidative Stability of Milk. *J. Dairy Sci.* 87: 406 - 12.

Alothman, M., Hogan, S. A., Hennessy, D. P., Dillon, K. N., Kilcawley, M., O'Donovan, J., Tobin, M., Fenelon, A. and O'Callaghan, T. F. 2019. The "Grass-Fed" Milk Story: Understanding the Impact of Pasture Feeding on the Composition and Quality of Bovine Milk. *Foods (Basel, Switzerland)* 8: 350.

Auerbach, Oldenburg, H., E. and Weissbach, F. 1998. Incidence of Penicillium Roqueforti and Roquefortine C in Silages. Article. *J. Sci. Food Agric.* 76: 565 - 72.

Austin, M. P. 2002. Spatial Prediction of Species Distribution: An Interface between Ecological Theory and Statistical Modelling. *Ecol. Modell.* 157: 101-18.

Bahr, C., Bruininx, E. and van den Belt, A. 2016. Nutrition, Sensors and Genetics – a Data Driven Smart Feeding Approach by Agrifirm. *Precision Dairy Farming*. 49: 53.

Banks, W., Clapperton, J. L. and Steele, W. 1983. Dietary Manipulation of the Content and Fatty Acid Composition of Milk-Fat. *Proc. Nutr. Soc.* 42: 399 - 406.

Beever, D.E., Gill, M. and Sutton, J.D. 1989. Limits to Animal Production with High Forage Diets. *J. Anim. Sci.* 67:298.

Beever, D. E., and Mould, F. L. 2000. Forage Evaluation for Efficient Ruminant Livestock Production. Chap. 2 In *Forage Evaluation in Ruminant Nutrition*, edited by Owen, E., Givens, D.I., Axford, R.F.E. and Omed, H.M. Wallingford Oxon OX10 8DE UK and Suite 3203, New York, NY. 10016 USA: CABI.

Bernardes, T. F., Daniel, J. L. P., Adesogan, A. T., McAllister, T. A., Drouin, P., Nussio, L. G., Huhtanen, P., Tremblay, G. F., Bélanger, G. and Cai, Y. 2018. Silage Review: Unique Challenges of Silages Made in Hot and Cold Regions. *J. Dairy Sci.* 101: 4001 – 4019.

Benchaar, C., Hassanat, F., Gervais, R., Chouinard, P. Y, Petit, H. V. and Massé, D. I.

2014. Methane Production, Digestion, Ruminal Fermentation, Nitrogen Balance, and Milk Production of Cows Fed Corn Silage- or Barley Silage-Based Diets. *J. Dairy Sci.* 97: 961-74.

Berckmans, D. and Keita, A. 2017. Precision Livestock Farming '17 *8th European Conference on Precision Livestock Farming*, Nantes, France.

Bertilsson, J., and Murphy, M. 2003. Effects of Feeding Clover Silages on Feed Intake, Milk Production and Digestion in Dairy Cows. *Grass Forage Sci.* 58: 309 - 22.

Blaney, B. J., McKenzie, R. A. Walters, J. R. Taylor, L. F. Bewg, W. S., Ryley, M. J. and Maryam, R. 2000. Sorghum Ergot (Claviceps Africana) Associated with Agalactia and Feed Refusal in Pigs and Dairy Cattle. *Aust. Vet. J.* 78: 102 - 7.

Borchani, H., Varando, G., Bielza, C. and Larranaga, P. 2015. A Survey on Multi - Output Regression. Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery 5: 216 - 33.

Brotzman, R. L., Cook, N. B. Nordlund, K. T. Bennett, Gomez, R. B. A. and Dopfer, D. 2015. Cluster Analysis of Dairy Herd Improvement Data to Discover Trends in Performance Characteristics in Large Upper Midwest Dairy Herds. *J. Dairy Sci.* 98: 3059 - 70.

Brown, L. D., Hillman, D., Lassiter, C. A. and Huffman, C. F. 1963. Grass Silage vs. Hay for Lactating Dairy Cows. *J. Dairy Sci.* 46: 407-10.

Buchanan and Smith, J. G. 2010. An Investigation into Palatability as a Factor Responsible for Reduced Intake of Silage by Sheep. J. *Anim. Sci.* 502: 253 - 60.

Cabrera, V. E., Barrientos, J. A., Blanco, H., Delgado and Fadul, L. and Pacheco. 2020. Symposium Review: Real - Time Continuous Decision Making Using Big Data on Dairy Farms *J. Dairy Sci.* 103: 3856 - 66.

Casper, D. P. and Schingoethe, D. J. 1986. Evaluation of Urea and Dried Whey in Diets of Cows During Early Lactation. Comparative Study. *J. Dairy Sci.* 69: 1346 - 54.

Chalupa, W., and Schneider, P. L. 1985. Buffers for Dairy Cattle. Paper presented at the 20th Annual Proceedings of the 1985 Pacific Northwest Animal Nutrition Conference, Boise, Idaho.

Canadian Dairy Information Centre 2015. Average Milk Production by Breed (Milk Recording). Accessed February 13, 2021. https://agriculture.canada.ca/en/canadas-agriculture-sectors/animal-industry/canadian-dairy-information-centre/dairy-statistics-and-market-information/dairy-animal-genetics/average-milk-production-breed

Christie, W. W. 1979. The Effects of Diet and Other Factors on the Lipid Composition of Ruminant Tissues and Milk. *Prog. Lipid Res* 17: 245 - 77.

Coblentz W. K., Hoffman P. C. 2009. Effects of Spontaneous Heating on Fiber Composition, Fiber Digestibility, and In Situ Disappearance Kinetics of Neutral Detergent Fiber for Alfalfa-Orchardgrass Hays. *J. Dairy Sci.*, 92:2875-95.

Combs, D. 2015. Forage Quality and Utilization: Total Tract NDF Digestibility. Western Alfalfa and Forage Symposium, University of Nevada, Reno, NV.

Coppock, R. W., Mostrom, M. S. Sparling, C. G. Jacobsen, B. and Ross, S. C. 1990. Apparent Zearalenone Intoxication in a Dairy Herd from Feeding Spoiled Acid - Treated Corn. Review. *Vet. Hum. Toxicol.* 32: 246-8.

Corbett, B. 1996. Heat Damage in Silage and Hay. Adapted from Alberta Agriculture Beef Herd Management Reference Binder and Study Guide – 307. Accessed January 1, 2021. www.Agriv.ov.ab.ca/crops/forage/silage

Dagenais, T. R., and Keller, N. P. 2009. Pathogenesis of Aspergillus Fumigatus in Invasive Aspergillosis. *Clin. Microbiol. Rev.* 22: 447 - 65.

Dairy Farmers of Canada 2019. Who We Are. 2019, Accessed October 15, 2020.

Danner, H., Hoizer, M., Weidenhoizer, E. and Braun, R. 2003. Acetic Acid Increases Stability of Silage under Aerobic Conditions. *Applied and Environmental Microbiology*, 69: 562 – 7.

Das, R., Sailo, L. Verma, N. Bharti, P. Saikia, J., Imtiwati and Kumar, R. 2016. Impact of Heat Stress on Health and Performance of Dairy Animals: A Review *Vet. World* 9: 260 - 8.

De Saeger, S., Sibanda, L., and Van Peteghem, C. 2003. Analysis of Zearalenone and A - Zearalenol in Animal Feed Using High-Performance Liquid Chromatography. *Anal. Chim. Acta* 487: 137 - 43.

Dean, A. and Voss, D. 1999. Design and Analysis of Experiments. Ed. Daniel Voss Angela Dean.

De'ath, G. 2002. Multivariate Regression Trees: A New Technique for Modeling Species-Environment Relationships. *Ecology* 83: 1105 - 17.

DePeters, E. J. and Taylor, S. J. 1985. Effects of Feeding Corn or Barley on Composition of Milk and Diet Digestibility. Research Support, Non-U.S. Gov't. *J. Dairy Sci.* 68: 2027 - 32.

Desmasures, N., Bazin, F. and Gueguen, M. 1997. Microbiological Composition of Raw Milk from Selected Farms in the Camembert Region of Normandy.

Dewhurst, R. J., Fisher, W. J., Tweed, J. K. S. and Wilkins, R. J. 2003. Comparison of Grass and Legume Silages for Milk Production. 1. Production Responses with Different Levels of Concentrate. *J. Dairy Sci.* 86: 2598-611.

Dey, P., and Pramit, P. 2020. Relevance of Data Transformation Techniques in Weed Science. *J. Res. Weed Sci.* 3: 81 - 89.

Dias, K.M., Garcia, S.G. and Clark, C.E.F. 2016. Creating Value of Data: Milking Order and Its Role in Future Precision Dairy Feeding Systems. *Precision Dairy Farming* 3: 383 – 85.

Driehuis, F., Wilkinson, J. M., Jiang, Y., Ogunade, I. and Adesogan, A. T. 2018. Silage Review: Animal and Human Health Risks from Silage. *J. Dairy Sci.* 101: 4093 - 110.

Duncan, A. J., and Milne, J. A. 1993. Effects of Oral Administration of Brassica Secondary Metabolites, Allyl Cyanide, Allyl Isothiocyanate and Dimethyl Disulphide, on the Voluntary Food Intake and Metabolism of Sheep. *Br. J. Nutr.* 7: 631 - 45. The Nutrition Society.

Environment Canada. Meteorological Service of Canada. Canadian Climate Normals. 1981-2010 Climate Normals & Averages. Accessed January 1, 2021. https://climate.weather.gc.ca/climate_normals/index_e.html

Eppard, P. J., Bauman, D. E., Bitman, J. D., Wood, L., Akers, R. M. and House, W. A. 1985. Effect of Dose of Bovine Growth Hormone on Milk Composition: Alpha - Lactalbumin, Fatty Acids, and Mineral Elements. *J. Dairy Sci.* 68: 3047 - 54.

Erdogan, H. M. 2010. Listerial Keratoconjunctivitis and Uveitis (Silage Eye). *Vet. Clin. North Am. Food Anim. Pract.* 26: 505 - 10.

Fatehi, F., Zali, A., Honarvar, M., Dehghan-banadaky, Young, M., A. J., Ghiasvand, M., Eftekhari, M. 2012. Review of the relationship between milk urea nitrogen and days in milk, parity, and monthly temperature mean in Iranian Holstein cows. J. Dairy Sci. 95: 5156 – 5163.

Review of the relationship between milk urea nitrogen and days in milk, parity, and monthly temperature mean in Iranian Holstein cows. J. Dairy Sci. 95: 5156 - 5163

Fenlon, D. R. 1986. Growth of Naturally-Occurring Listeria Spp in Silage - a Comparative-Study of Laboratory and Farm Ensiled Grass. *Grass Forage Sci.* 41: 375 - 78.

Fenlon, D.R. 1988. Listeriosis Review of B.A. Stark, J.M. Wilkinson. Eds. *Silage and Health.* 3: 7 - 18.

Fenlon, D. R., Wilson, J. and Weddell, J. R. 1989. The Relationship between Spoilage and Listeria-Monocytogenes Contamination in Bagged and Wrapped Big Bale Silage. *Grass Forage Sci.* 44: 97 – 100.

Forbes, J. M., Mbanya, J. N. and Anil, M. H. 1992. Effects of Intraruminal Infusions of Sodium Acetate and Sodium Chloride on Silage Intake by Lactating Cows. *Appetite* 19: 293 - 301.

Gallo, A., Bertuzzi, T. Giuberti, G. Moschini, M. Bruschi, S. Cerioli, C. and Masoero, F. 2016. New Assessment Based on the Use of Principal Factor Analysis to Investigate Corn Silage Quality from Nutritional Traits, Fermentation End Products and Mycotoxins. *J. Sci. Food Agric.* 9: 437 - 48.

Gandhi, M., and Chikindas, M. L. 2007. Listeria: A Foodborne Pathogen That Knows How to Survive. *Int. J. Food Microbiol.* 113: 1 - 15.

Garcia, A. D., Olson, W. G., Otterby, D. E., Linn, J. G., Hansen, W. P. 1989. Effects of Temperature, Moisture, and Aeration on Fermentation of Alfalfa Silage. *J. Dairy Sci*. 72:93-103.

Gerlach, K., Luiz, J. P. D., Cabreira, C. J. and Luiz, G. N. 2020. A Data Analysis on the Effect of Acetic Acid on Dry Matter Intake in Dairy Cattle. *Anim. Feed Sci. Technol.* 272: 114782.

Gianola, D., Okut, H. Weigel, K. A. and Rosa, G. J. 2011. Predicting Complex Quantitative Traits with Bayesian Neural Networks: A Case Study with Jersey Cows and Wheat. *BMC Genet* 12: 87.

Grant, R. J., and Ferraretto, L. F. 2018. Silage Review: Silage Feeding Management: Silage Characteristics and Dairy Cow Feeding Behavior. *J. Dairy Sci.* 101: 4111 - 21.

Grayson, R. R. 1956. Silage Gas Poisoning: Nitrogen Dioxide Pneumonia, a New Disease in Agricultural Workers. Article. *Ann Intern Med.* 45: 393 - 408.

Guan, H., Shuai, Y., Yan, Y., Ran, Q., Wang, X., Li, D., Cai, Y., Zhang, X. 2020. Microbial Community and Fermentation Dynamics of Corn Silage Prepared with Heat-Resistant Lactic Acid Bacteria in a Hot Environment. Microorganisms. 8: 719

Gupta, S.C. and Kapoor., V.K. 1990. Fundamentals of Mathematical Statistics. Ed. 2014.

Haggblom, P. Isolation of Roquefortine C from Feed Grain. Article. *Appl. Environ. Microbiol.* 56: 2924 - 6.

Hammami, H., Bormann, J., M'hamdi, N., Montaldo, H.H., Gengler, N. 2013. Evaluation of Heat Stress Effects on Production Traits and Somatic Cell Score of Holsteins in a Temperate Environment. *J. Dairy Sci.* 96:1844-1855

Harty, E., and Healy, J. 2016. Using Big Data and Advanced Analytics Top Optimize Health and Fertility. *In Precision Dairy Farming* 3: 219 – 22.

Havemose, M. S., Weisbjerg, M. R. Bredie, W. L. P. and Nielsen, J. H. 2004. Influence of Feeding Different Types of Roughage on the Oxidative Stability of Milk. *Int. Dairy J.* 14: 563 - 70

Ho, A. J., Ivanek, R. Grohn, Y. T., Nightingale, K. K. and Wiedmann, M. 2007. Listeria Monocytogenes Fecal Shedding in Dairy Cattle Shows High Levels of Day - to - Day Variation and Includes Outreaks and Sporadic Cases of Shedding of Specific L. Monocytogenes Subtypes. *Prev. Vet. Med.* 80: 287 - 305.

Hoedtke, S., Gabel, M. and Zeyner, A. 2010. Der Proteinabbau Im Futter Während Der Silierung Und Veränderungen Der Zusammensetzung Der Rohproteinfraktion. *Übers. Tierernährg.* 38: 157 - 79.

Hoffman, P., Combs, D. and Casler, M. 1998. Performance of Lactating Dairy Cows Fed Alfalfa Silage or Perennial Ryegrass Silage. *J. Dairy Sci.* 81: 162-8.

Hoffman, P. C., Shaver, R. D., Combs, D. K., Undersander, D. J., Bauman, L. M., Seeger, T. K. 2001. Understanding NDF Digestibility of Forages. *Wisconsin Team Forage*. 3:

Hoffman, P. C., Sievert, S.J., Shaver, R.D., Welch, D.A., Combs, D.K., 1993. In Situ Dry Matter, Protein, and Fiber Degradation of Perennial Forages. *J. Dairy Sci.*76:2632 - 2643

Huhtanen, P., Khalili, H., Nousiainen, Rinne, J. I., Jaakkola, M. S., Heikkila, T. and Nousiainen, J. 2002. Prediction of the Relative Intake Potential of Grass Silage by Dairy Cows. *Livest. Prod. Sci.* 73: 111 - 30.

Huhtanen, P., Nousiainen, J. I., Khalili, H., Jaakkola, S. and Heikkila, T. 2003. Relationships between Silage Fermentation Characteristics and Milk Production Parameters: Analyses of Literature Data. *Livest.Prod. Sci.* 81: 57 - 73.

Huhtanen, P., Rinne, M. and Nousiainen, J. 2007. Evaluation of the Factors Affecting Silage Intake of Dairy Cows: A Revision of the Relative Silage Dry - Matter Intake Index. *Anim.* 1: 758 - 70.

Huhtanen, P., Rinne, M. and Nousiainen, J. 2008. Evaluation of Concentrate Factors Affecting Silage Intake of Dairy Cows: A Development of the Relative Total Diet Intake Index. Consortium. *Anim.* 2: 942 - 53.

Jenness, R. 1985. Biochemical and Nutritional Aspects of Milk and Colostrum. Ed. Ames: Iowa State University Press.

Joint FAO / WHO Expert Committee on Food Additives. 2001. Evaluation of Certain Mycotoxins in Food. Fifty - Sixth Report of the Joint FAO / Who Expert Committee on Food Additives. Report. no. 0512 - 3054 (Print) (2001 - 2002).

Kalač, P. 2011. The Effects of Silage Feeding on Some Sensory and Health Attributes of Cow's Milk: A Review. *Food Chem.* 125: 307 - 17.

Kallela, K., Heinonen, K. and Saloniemi, H. 1984. Plant Oestrogens; the Cause of Decreased Fertility in Cows. A Case Report. *Nord Vet. Med.* 36: 124 - 9.

Kamphuis, C., and Steeneveld, W. 2016. *Precision Dairy Farming 2016.*

Kawas, J. R., Jorgensen, N. A., Hardie, A. R. and Danelon, J. L. 1983. Change in Feeding Value of Alfalfa Hay with Stage of Maturity and Concentrate Level. *J. Dairy Sci.* 66: 181

Kitchen, B. J. 1981. Review of the Progress of Dairy Science: Bovine Mastitis: Milk Compositional Changes and Related Diagnostic Tests. Review. *J. Dairy Res.* 48: 167 - 88.

Křížek, M. 2009. The Determination of Biogenic Amines in Silage. *Archiv für Tierernaehrung* 41: 97 - 104.

Kuiper, Goodman, T., Scott, P. M. and Watanabe, H. 1987. Risk Assessment of the Mycotoxin Zearalenone. Article. *Regul. Toxicol. Pharmacol.* 7: 253 - 306.

Kuoppala, K., Rinne, M., Nousiainen, J. and Huhtanen, P. 2008. The Effect of Cutting Time of Grass Silage in Primary Growth and Regrowth and the Interactions between Silage

Quality and Concentrate Level on Milk Production of Dairy Cows. *Livest. Prod. Sci.* 116: 171 - 82.

Lean, I. J. 2001. Association between Feeding Perennial Ryegrass (Lolium Perenne Cultivar Grasslands Impact) Containing High Concentrations of Ergovaline, and Health and Productivity in a Herd of Lactating Dairy Cows. *Aust. Vet. J.* 3: 262 - 4.

Leduc, M. 2019. Top-Quality Silage with Oxygen-Limiting Silos. Lactanet. Accessed February 10, 2022. https://www.valacta.com/en-ca/library/top-quality-silage-oxygen-limiting-silos

Linn, J. G. 1988. Factors Affecting the Composition of Milk from Dairy Cows. Designing Foods: Animal Product Options in the Marketplace.

Litherland, N. 2018. Improving Peak Milk Yields.

Litwińczuk, Z., Król, J., Brodziak, A., Barłowska, J. 2011. Changes of Protein Content and its Fractions in Bovine Milk from Different Breeds Subject to Somatic Cell Count. *J. Dairy Sci.* 94: 684 - 91.

Lokhorst, C., de Mol, R. M. and Kamphuis, C. 2019. Invited Review: Big Data in Precision Dairy Farming. *Anim.* 13: 1519 - 28.

Lokhorst, C., Kroeze, G.H., van der Berg, J.V. Blok, B. and de Vries, F. 1999. Datamining for Knowledge Discovery in Dutch Dairy Databases. Paper presented at the Second European Conference of the European Federation for lnformation on Perspectives of Modern Information and Communication Systems in Agriculture, Food Production and Environmental Technology in Agriculture, Food and the Environment, Universität Bonn-ILB, Bonn, Germany.

Lundquist, R. G., Linn, J. G. and Otterby, D. E. Influence of Dietary Energy and Protein on Yield and Composition of Milk from Cows Fed Methionine Hydroxy Analog. Comparative Study.

Research Support. 1983. Non - U.S. Gov't. *J. Dairy Sci.* 66: 475 - 91.

Marshall, J. C., Soboleva, T. K., Jamieson, P. and French., N. P. 2016. Estimating Bacterial Pathogen Levels in New Zealand Bulk Tank Milk. *J. Food Prot.* 79: 771 - 80.

Mathur, S., Constable, P. D., Eppley, R. M., Waggoner, M A. Tumbleson, L. E. and Haschek, W. M. 2001. Fumonisin B (1) Is Hepatotoxic and Nephrotoxic in Milk - Fed Calves. Article. *Toxicol. Sci.* 60: 385 - 96.

McLean, D. M., Graham, E. R., Ponzoni, R. W. and McKenzie, H. A. 1984. Effects of Milk-protein Genetic Variants on Milk Yield and Composition. *J.Dairy Res.* 51: 531 - 46.

Mertens, D. and Berzaghi, P. 2009. Silage Quality and Dairy Production. Paper presented at the International Silage Conference, Madison, WI.

Minson, D.J. 1990. Forage in Ruminant Nutrition. Ed Academic Press.

Mishra, D. 2019. Regression: An Explanation of Regression Metrics and What Can Go Wrong. Accessed March 5, 2021. https://towardsdatascience.com/regression-an-explanation-of-regression-metrics-and-what-can-go-wrong-a39a9793d914?gi=b2eac4ecc40f

Monteiro, H. F., Paula, E. M., Muck, R. E., Broderick, G. A. and Faciola, A. P. 2021. Effects of Lactic Acid Bacteria in a Silage Inoculant on Ruminal Nutrient Digestibility, Nitrogen Metabolism, and Lactation Performance of High-Producing Dairy Cows. *J. Dairy Sci.* 104: 8826 - 34.

Montgomery D. C., Peck, E. A. and Vining G. G. 2017. Introduction to Linear Regression Analysis*.*

Montgomery, D. C. 2013. Design and Analysis of Experiments.

Muck, R. E. 1988. Factors Influencing Silage Quality and Their Implications for Management. *J. Dairy Sci.* 71: 2992 - 3002.

Naude, T. W., Botha, C. J., Vorster, J. H., Roux, C. E., Van der Linde, J. S., Van der Walt, I. G., Rottinghaus, E., Van Jaarsveld, L. and Lawrence, A. N. 2005. Claviceps Cyperi, a New Cause of Severe Ergotism in Dairy Cattle Consuming Corn Silage and Teff Hay Contaminated with Ergotised Cyperus Esculentus (Nut Sedge) on the Highveld of South Africa. *Onderstepoort. J. Vet. Res.* 72: 23 - 37.

Nguyen, Q. T., Fouchereau, R., Frenod, E., Gerard, C. and Sincholle, V. 2020. Comparison of Forecast Models of Production of Dairy Cows Combining Animal and Diet Parameters. *Comput. Electron. Agric.* 170:

Nightingale, K. K., Schukken, Y. H., Nightingale, C. R., Fortes, E. D., Ho, A. J., Her, Z. Y., Grohn, T. P., McDonough, L. and Wiedmann, M. 2004. Ecology and Transmission of Listeria Monocytogenes Infecting Ruminants and in the Farm Environment. *Appl. Environ. Microbiol.* 70: 4458 - 67.

Nousiainen, J., Rinne, M. and Huhtanen, P. 2009. A Meta - Analysis of Feed Digestion in Dairy Cows. 1. The Effects of Forage and Concentrate Factors on Total Diet Digestibility. Interpretable Summary. *J. Dairy Sci.* 92: 5019 - 30.

Ocak, E. and Inci, S. 2008. The Effect of Seasonal Variation on the Composition of Cow Milk in Van Province. *Pakistan JN.* 7

Oetzel, G. 2007. Herd - Level Ketosis – Diagnosis and Risk Factors.

Ogunade, I. M., Martinez-Tuppia, C., Queiroz, O. C. M., Jiang, Y., Drouin, P., Wu, F., Vyas, D. and Adesogan, A. T. 2018. Silage Review: Mycotoxins in Silage: Occurrence, Effects, Prevention, and Mitigation. *J. Dairy Sci.* 101: 4034 - 59.

O'Kiely, P., Turley, T. and Rogers, P. A. 1999. Exposure of Calves to Nitrogen Dioxide in Silage Gas. *Vet. Rec.* 144: 352 - 3.

Oliveira, A. S., Weinberg, Z. G., Cervantes, A. A., Arriola, P. K. G., Ogunade, I. M., Jiang, Y., Kim. 2016. 0636 Meta-Analysis of the Effect Silage Inoculation with Homolactic or Facultative Heterolactic Bacteria on the Performance of Dairy Cows. *J. Anim. Sci.* 94: 303-04.

O'Mara, F. P., Fitzgerald, J. J., Murphy, J. J. and Rath, M. 1998. The Effect on Milk Production of Replacing Grass Silage with Corn Silage in the Diet of Dairy Cows. *Livest. Prod. Sci.* 55: 79 - 87.

Oney, C. R., Gramkow, J. K., Hilscher, F. H., Watson, A. K., Erickson, G. E., MacDonald, J. C., Klopfenstein, T. J. 2018. Rumen Undegradable Protein Content and Digestibility of Corn Silage and High-moisture Corn. *Nebraska Beef Cattle Report.* pg. 52-53.

Oscar, N. D. M., Aello, M. S., Nomdedeu, M. Houtte, S. V. 2002. Effect of Maize Crop Maturity on Silage Chemical Composition and Digestibility (In Vivo, In Situ and In Vitro). *Anim. Feed Sci. Technol.* 99: 37 - 43.

Osweiler, G. D., Kehrli, M. E., Stabel, J. R., Thurston, J. R., Ross, P. F. and Wilson, T. M. 1993. Effects of Fumonisin - Contaminated Corn Screenings on Growth and Health of Feeder Calves. *J. Anim. Sci.* 71: 459 - 66.

Pang, D. G., Krizsan, S. J., Sairanen, Nousiainen, A. J. and Huhtanen, P. 2019. Modelling Feed Intake and Milk Yield Responses to Different Grass Ley Harvesting Strategies. *Grass Forage Sci.* 74: 509 – 24.

Patel, M. 2012. Effects of Increasing the Proportion of High - Quality Grass Silage in the Diet of Dairy Cows. Doctoral thesis, Acta Universitatis Agriculturae Sueciae.

Phillips, D. 2017. Canada’s Top Ten Weather Stories 2017. Environment and Climate Change Canada. Accessed June 21, 2021. https://www.canada.ca/en/environment-climate-change/services/top-ten-weather-stories/2017.html

Producteurs De Lait Du Québec. 2019. Profile and Impact of Milk Economy. Accessed August 3. https://lait.org/en/the-milk-economy/profile-and-impact-of-milk-economy/

Provenza, F. D., Burritt, E. A., Clausen, T. P., Bryant, J. P., Reichardt, P. B. and Distel, R. A. 1990. Conditioned Flavor Aversion - a Mechanism for Goats to Avoid Condensed Tannins in Blackbrush. *Am. Nat.* 136: 810 - 28.

Puntenney, S., Wang, Y. and Forsberg, N. 2020. Keeping Them out of the Rough: Practical Insights into Hemorrhagic Bowel Syndrome. *Toxicol. Sci*.

Rachid., B., Lahmar, M., Majdoub, A., Djemali, M. and Belyea, R. 2002. The Relationship of Temperature - Humidity Index with Milk Production of Dairy Cows in a Mediterranean Climate. *Anim. Res.* 51: 479-91.

Rangaswami, R. A. 2018. Textbook of Agricultural Statistics. New Age International Publishers.

Reu, K., Grijspeerdt, K. and Herman, L. 2004. A Belgian Survey of Hygiene Indicator Bacteria and Pathogenic Bacteria in Raw Milk and Direct Marketing of Raw Milk Farm Products. *J. Food Saf.* 24: 17 - 36.

Ribeiro - Oliveira, J. P., Santana, D. G., de Pereira, V. J. and dos Santos, C. M. 2018. Data Transformation: An Underestimated Tool by Inappropriate Use. *Acta Sci. Agron.* 40:

Roca - Fernández, A. I. and González – Rodríguez, A. 2013. Sward Factors Influence on Pasture Dry Matter Intake of Grazing Dairy Cows: Review Article. *Iran. J. Appl. Anim. Sci.* 3: 624 - 51.

Rodrigues, I. and Naehrer, K. 2012. A Three - Year Survey on the Worldwide Occurrence of Mycotoxins in Feedstuffs and Feed. Article. *Toxins (Basel)* 4: 663 - 75.

Rolleri, G. D., Larson, B. L. and Touchberry, R. W. 1956. Protein Production in the Bovine - Breed and Individual Variations in the Specific Protein Constituents of Milk. *J. Dairy Sci.* 39: 1683 - 89.

Romney, D.L. and Gill, M. 2000. Intake of Forages. Chap. 3 In *Forage Evaluation in Ruminent Nutrition* edited by E. Owen D.I. Givens, R.F.E. Axford and H.M. Omed. Wallingford Oxon OX10 8DE UK and Suite 3203 New York, NY. 10016 USA.: CABI.

Ryser, E. T., Arimi, S. M. and Donnelly, C. W. 1997. Effects of Ph on Distribution of Listeria Ribotypes in Corn, Hay, and Grass Silage. *Appl. Environ. Microbiol.* 63: 3695 - 7.

Sanaa, M., Poutrel, B. Menard, J. L. and Serieys, F. 1993.Risk Factors Associated with Contamination of Raw Milk by Listeria Monocytogenes in Dairy Farms. *J. Dairy Sci.* 76: 2891 - 8.

Sanchez - Fernandez, M., de - Prado - Cumplido, M., Arenas - Garcia, J. and Perez – Cruz, F. 2004. SVM. Multi-Regression for Nonlinear Channel Estimation in Multiple - Input Multiple - Output Systems. *IEEE. Trans. Signal* 52: 2298 - 307.

Sarfati, J., Jensen, H. E. and Latge, J. P. 1996. Route of Infections in Bovine Aspergillosis. *J. Med. Vet. Mycol.* 34: 379 - 83.

Scherer, R., Gerlach, K. and Sudekum, K. H. 2015. Biogenic Amines and Gamma - Amino Butyric Acid in Silages: Formation, Occurrence and Influence on Dry Matter Intake and Ruminant Production. *Anim. Feed Sci. Technol.* 210: 1 - 16.

Schick, B., 2019. Baleage vs. Haylage: What’s Really the Difference? Institute of Agriculture and Natural Resources.

Schneider, D. J., Miles, C. O., Garthwaite, I. A., Van Halderen, Wessels, J. C. and Lategan, H. J. 1996. First Report of Field Outbreaks of Ergot - Alkaloid Toxicity in South Africa. Article. *Onderstepoort J. Vet. Res.* 63: 97 - 108.

Seglar, B. 2003. Fermentation Analysis and Silage Quality Testing. Minnesota Dairy Health Conference, St. Paul Minnesota.

Shaver, R. D., Nytes, A. J., Satter, L. D. and Jorgensen, N. A. 1986. Influence of Amount of Feed-Intake and Forage Physical Form on Digestion and Passage of Prebloom Alfalfa Hay in Dairy - Cows. *J. Dairy Sci.* 69: 1545 - 59.

Smith, A. M. 1954. Seasonal Variation in the Quality of Grass Silage. *J. Sci. Food Agric.* 5: 48 - 51.

Smith, D. F., and Lynch, G. P. 1973. Aspergillus Fumigatus in Samples of Moldy Silage. *J. Dairy Sci.* 56: 828 - 9.

Solórzano, L. C., Solórzano, L. L., Rodríguez-Carías, A. A., Solórzano, L. C., Solórzano, L. L., and Rodríguez, A. A. 2020. Models of Laboratory Mino-Silos to Study the Fermentation of Wilted Alfalfa (Medicago Sativa) Using a Homolactic Bacterial Inoculant. *The Journal of Agriculture of the University of Puerto Rico*, *104*: 223 – 232.

Spyromitros - Xioufis, E., Groves, W., Tsoumakas, G. and Vlahavas, I. 2012. Multi - Label Classification Methods for Multi - Target Regression.

Südekum, K.-H. and Eisner, I. 2009. Silagen: Voraussetzungen Für Eine Hohe Akzeptanz Und Maximale Trockenmasse - Aufnahme. *Übers. Tierernährg.* 37: 133 - 40.

Sutton, J. D. 1980. Influence of Nutritional Factors on the Yield and Content of Milk-Fat: Dietary Components Other Than Fat. *Int. Dairy Fed. Bull. Doc.* 3: 125 - 26.

Sutton, J. D. 1985. Digestion and Absorption of Energy Substrates in the Lactating Cow. *J. Dairy Sci.* 68: 3376 - 93.

Tedeschi, L. O. 2019. Asn - Asas Symposium: Future of Data Analytics in Nutrition: Mathematical Modeling in Ruminant Nutrition: Approaches and Paradigms, Extant Models, and Thoughts for Upcoming Predictive Analytics 1, 2. *J. Anim. Sci.* 97: 1921 - 44.

Terrence, V. and Michele, V. 2009. The Canadian Encyclopedia. Agriculture and Food, Historica Canada. Accessed Feb. 11, 2021. https://www.thecanadianencyclopedia.ca/en/article/agriculture-and-food

Thomas, C., Aston, K. and Daley, S. R. 1985. Milk Production from Silage 3. A Comparison of Red Clover with Grass Silage. *J. Anim. Sci.* 41: 23-31.

The Minitab Blog Editor 2017. What Is the Difference between Linear and Nonlinear Equations in Regression Analysis? Accessed November 30, 2020. https://blog.minitab.com/en/adventures-in-statistics-2/what-is-the-difference-between-linear-and-nonlinear-equations-in-regression-analysis#:~:text=While%20a%20linear%20equation%20has,the%20term%20%E2%80%9Cnonlinear%E2%80%9D%20itself.&text=Unlike%20linear%20regression%2C%20these%20functions,one%20parameter%20per%20predictor%20variable.

Unnerstad, H., Romell, A. H., Ericsson, M. L., Danielsson-Tham, and Tham, W. 2000. Listeria Monocytogenes in Faeces from Clinically Healthy Dairy Cows in Sweden. *Acta Vet. Scand.* 41: 167 - 71.

Van der Heide, E. M. M., Veerkamp, R. F., van Pelt, M. L., Kamphuis, C., Athanasiadis, I. and Ducro, B. J. 2019. Comparing Regression, Naive Bayes, and Random Forest Methods in the Prediction of Individual Survival to Second Lactation in Holstein Cattle. *J. Dairy Sci.* 102: 9409 - 21.

Van der Veen, S., Moezelaar, R., Abee, T. and Wells-Bennik, M. H. 2008. The Growth Limits of a Large Number of Listeria Monocytogenes Strains at Combinations of Stresses Show Serotype -- and Niche - Specific Traits. *J. Appl. Microbiol.* 105: 1246 - 58.

Van der Waaij, B. D., Feldbrugge, R. L. and Veerkamp, R. F. 2016. Cow Feed Intake Prediction with Machine Learning. *In Precision Dairy Farming* 3: 377 – 83.

Van Kessel, J. A., Karns, J. S., Lombard, J. E. and Kopral, C. A. 2011. Prevalence of Salmonella Enterica, Listeria Monocytogenes, and Escherichia Coli Virulence Factors in Bulk Tank Milk and in - Line Filters from U.S. Dairies. *J. Food. Prot.* 74: 759 - 68.

Van Soest, P. J. 1994. Nutritional Ecology of the Ruminant. Ithaca: Cornell University Press.

Vasquez, E., and Walter, E. 2003. Multi - Output Support Vector Regression. Proceedings of the Thirteen IFAC Symposium on System Identification, Rotterdam, The Netherlands.

Verhoosel, J. P. C., and Spek, J. 2016. Semantics for Big Data Applications in the Smart Dairy Farming Domain. Paper presented at the 1st Precision Dairy Farming.

Vilar, M. J., Yus, E., Sanjuan, M. L., Dieguez, F. J. and Rodriguez – Otero, J. L. 2007. Prevalence of and Risk Factors for Listeria Species on Dairy Farms. *J. Dairy Sci.* 90: 5083 - 8.

Vlaeminck, B., Fievez, V., Tamminga, S., Dewhurst, R. J., van Vuuren, A., De Brabander, D. and Demeyer, D. 2006. Milk Odd- and Branched-Chain Fatty Acids in Relation to the Rumen Fermentation Pattern. *J. Dairy Sci.* 89: 3954 - 64.

Warton, D. I., and Hui, F. K. 2011. The Arcsine Is Asinine: The Analysis of Proportions in Ecology. *Ecology* 92: 3 - 10.

Wiedmann, M., Arvik, T., Bruce, J. L., Neubauer, del Piero, J. F., Smith, M. C., Hurley, J., Mohammed, H. O. and Batt, C. A. 1997. Investigation of a Listeriosis Epizootic in Sheep in New York State. *Am. J. Vet. Res.* 58: 733-7.

Woodford, J. A., Jorgensen, N. A. and Barrington, G. P. 1986. Impact of Dietary Fiber and Physical Form on Performance of Lactating Dairy Cows. Comparative Study Research Support, Non - U.S. Gov't. *J. Dairy Sci.* 69: 1035 - 47.

Yoon, J. T., Lee, J., Kim, C., Chung, Y. and Kim, C. 2004. Effects of Milk Production, Season, Parity and Lactation Period on Variations of Milk Urea Nitrogen Concentration and Milk Components of Holstein Dairy Cows. *Asian-Australalasian J. Anim. Sci*.

Yu, Y. 1976. Relationship Between Measurements of Heating and Acid-Detergent Insoluble Nitrogen in Heat Damaged Fresh Alfalfa, Haylage, and Hay. *J. Dairy Sci.*

Zain, M.E. 2011. Impact of Mycotoxins on Humans and Animals. Review. *J. Saudi Chem. Soc.* 15: 129 - 44.

Zar, J. 1999. Biostatistical Analysis.

Zhang, W., Xianhui, L., Yi, D., Deming, S. 2012. Multi - Output LS-SVR Machine in Extended Feature Space. Paper presented at the 2012 IEEE. International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings.

Zhao, Y., Nasrullah, Z. and Li, Z. 2019. Pyod: A Python Toolbox for Scalable Outlier Detection. *J. Mach. Learn. Res.* 20: 96:1-96:7.

Zhong, Y., Xue, M. and Liu, J. 2018. Composition of Rumen Bacterial Community in Dairy Cows with Different Levels of Somatic Cell Counts. *Front. Microbiol.* 9: 3217.

APPENDICES

Appendix 1: General Information About Variables Used for Silage Variable Prediction and Mixed Effects Model

Variable	Unit	Grass and Legume Silage	Corn Silage	Milk Production
1, 2 – Propanediol	DM %		X	
Acetic Acid	DM %	X	X	
Acid Detergent Fiber (ADF)	NDF %	X	X	X
Acid Detergent Fiber	DM %	X	X	X
Acid Detergent Insoluble Crude Protein (ADICP)	CP %	X	X	X
Acid Detergent Insoluble Crude Protein	DM %	X	X	X
Ash	DM %	X	X	X
Average Daily Fat Per Cow	Kg			X
Average Daily Milk Per Cow	Kg			X
Average Daily Protein Per Cow	Kg			X
Average Days in Milk	Number of days			X
Average Somatic Cell Count	Somatic cell count average of 1000 cells per mL			X
Butyric Acid	DM %	X		
Crude Fat	DM %	X	X	X
Crude Protein (CP)	DM %	X	X	X
Dry Matter (DM)	DM %	X	X	X
Herd	ID	X	X	X
Herd Test Period	ID	X	X	X
Lactic acid	DM %	X	X	
Lignin	DM %	X	X	X
Lignin	NDF %	X	X	X
Milk Urea Nitrogen	% In milk-protein			X
Neutral Detergent Fiber (NDF)	DM %	X	X	X
Neutral Detergent Insoluble Crude Protein (NDICP)	CP %	X	X	X
Neutral Detergent Insoluble Crude Protein	DM %	X	X	X
Nonfibrous Carbohydrate (NFC)	DM %	X	X	X
pH	N/A	X	X	X
Propionic Acid	DM %	X	X	
Proportion of Concentrate in Feed	DM %	X	X	
Proportion of Other Silage in Feed	DM %	X	X	
Region	Region			X
Rumen Digestible Neutral Detergent Fiber at 30 Hour in Vitro (NDFD30)	NDF %	X	X	
Rumen Digestible Neutral Detergent Fiber at 120 Hour in Vitro (NDFD120)	NDF %	X	X	
Rumen Digestible Neutral Detergent Fiber at 240 Hour in Vitro (NDFD240)	NDF %	X	X	
Soluble Protein (SP)	CP %	X	X	X
Soluble Protein	DM %	X	X	X
Starch	DM %		X	X
Starch Digestibility (StarchD)	DM %		X	X
Storage Method	Storage Method			X
Test Month	Month			X
Test Year	Year			X

N/A: Not Applicable

Appendix 2: Feature and Target Variables for Silage Variables Imputation

Variable	Type	Grass and Legume Silage	Corn Silage
1, 2 – Propanediol	Target		X
Acetic Acid	Target	X	X
Acid Detergent Fiber (ADF)	Feature	X	X
Acid Detergent Fiber	Feature	X	X
Acid Detergent Insoluble Crude Protein (ADICP)	Feature	X	X
Acid Detergent Insoluble Crude Protein	Feature	X	X
Ash	Feature	X	X
Butyric Acid	Target	X	
Crude Fat	Feature	X	X
Crude Protein (CP)	Feature	X	X
Dry Matter (DM)	Feature	X	X
Lactic acid	Target	X	X
Lignin	Feature	X	X
Lignin	Feature	X	X
Neutral Detergent Fiber (NDF)	Feature	X	X
Neutral Detergent Insoluble Crude Protein (NDICP)	Feature	X	X
Neutral Detergent Insoluble Crude Protein	Feature	X	X
Nonfibrous Carbohydrate (NFC)	Feature	X	X
pH	Target	X	X
Propionic Acid	Target	X	X
Rumen Digestible Neutral Detergent Fiber at 30 Hour in Vitro (NDFD30)	Target	X	X
Rumen Digestible Neutral Detergent Fiber at 120 Hour in Vitro (NDFD120)	Target	X	X
Rumen Digestible Neutral Detergent Fiber at 240 Hour in Vitro (NDFD240)	Target	X	X
Soluble Protein (SP)	Feature	X	X
Soluble Protein	Feature	X	X
Starch	Feature		X
Starch Digestibility (StarchD)	Feature		X