Machine Learning for Accurate Cost Estimation in Underground Mining: A Contractor's Perspective

Juan Camilo García Vásquez

Department of Mining and Materials Engineering

McGill University, Montreal

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

Abstract

Accurate cost estimation is crucial in effective decision-making and evaluation in underground mining projects. Machine learning techniques have shown enormous potential in enhancing the accuracy of cost estimation in various industries. This study harnesses artificial neural networks (ANN) and Support Vector Machines (SVM) to estimate operating costs in underground mining. Special emphasis is placed on cost estimation from a contractor's perspective.

Mining contractors are sensitive to deviations from the estimated costs because slight deviations may result in losing a contract bid or economic loss in an awarded project. The proposed approach can help contractors make more informed decisions and improve project management. Comprehensive data containing various parameters that impact the cost of underground mining projects, such as equipment type utilization, rock type, and cross-sectional area, were collected. This dataset was used to train and evaluate ANN and SVR models that can significantly improve the accuracy of cost estimation in underground mining. The best model achieved a mean average percentage error (MAPE) of 5.31% for the ANN model and 3.05% for the SVR one, outperforming traditional cost estimation methods. This study demonstrates the potential of machine learning in enhancing the efficiency and accuracy of cost estimation in underground mining projects.

Abstrait

L'estimation précise des coûts est cruciale dans la prise de décision et l'évaluation efficace des projets miniers souterrains. Les techniques d'apprentissage automatique ont montré un énorme potentiel pour améliorer la précision de l'estimation des coûts dans diverses industries. Cette étude exploite les réseaux neuronaux artificiels (RNA) et les machines à vecteurs de support (MVS) pour estimer les coûts d'exploitation dans les mines souterraines. Un accent particulier est mis sur l'estimation des coûts du point de vue d'un entrepreneur.

Les entrepreneurs miniers sont sensibles aux écarts par rapport aux coûts estimés, car de légères variantes peuvent entraîner la perte d'une soumission de contrat ou des pertes financières dans un projet attribué. L'approche proposée peut aider les entrepreneurs à prendre des décisions plus éclairées et à améliorer la gestion de projet. Des données complètes contenant divers paramètres ayant un impact sur le coût des projets miniers souterrains, tels que l'utilisation du type d'équipement, le type de roche et la section transversale, ont été collectés. Ce jeu de données a été utilisé pour former et évaluer les modèles de RNA et de MVS qui peuvent améliorer considérablement la précision de l'estimation des coûts dans les mines souterraines. Le meilleur modèle a atteint un pourcentage d'erreur absolu moye (MAPE) de 5,31 % pour le modèle de RNA et 3,05 % pour celui de MVS, surpassant les méthodes traditionnelles d'estimation des coûts. Cette étude nous démontre le plein potentiel de l'apprentissage automatique afin d'améliorer l'efficacité et la précision de l'estimation des coûts dans les projets miniers souterrains.

Acknowledgements

I would like to thank Professor Mustafa Kumral for his invaluable guidance, support and understanding during my master´s degree.

Thank you to all the people who have accompanied me during the years of education, those who have left, those who are still here and those who will come.

Thanks to all my colleagues in the Reliability Analysis and Maintenance Lab for their patience and support during this time, including Mehmet, Zhixuan, Zhanbolat, Sena, Carolina, Arturo, Eduardo and especially Max, who was patient enough to support me during my introduction to this world of programming and machine learning and gave me great ideas for the implementation of these techniques.

I want to thank my family, especially to my grandmother, my aunts Elvia and Martha, and my uncle Fredy, since without their help, I would not have been able to establish the foundations to be where I am today.

I also want to thank my friends, who, despite the distance, were always present, especially in the most challenging times. After all, our friends are the family we chose.

Thanks also to MITACS and Dumas Mining for supporting this research (MITACS IT32728 ACC).

Contributions of Authors

The author of this thesis is Juan Camilo García Vásquez.

Professor Mustafa Kumral was the supervisor of the author´s Master of Science Degree, and co-authored article titled as “Artificial Neural Networks for Accurate Cost Estimation in Underground Mining: A Contractor's Perspective,” which has been submitted.

Table of Contents

List of figures

List of tables

List of abbreviations

ANN	Artificial Neural Network
AACE	Association for the Advancement of Cost Engineering
ANSI	American National Standards Institute
MAE	Mean Absolute Error
MAPE	Mean Average Percentage Error
MSE	Mean Square Error
SVM	Support Vector Machine
PMBOK	Project Management Body of Knowledge
PMIS	Project Management Information System
RFP	Request for Proposal
RMSE	Root Mean Square Error
WBS	Work Breakdown Structure

1 Introduction

1.1 Objectives

Cost estimation is a specific challenge for mining contractors. Mining contractors are highly sensitive to cost estimation accuracy because they are relatively small-scale corporations that provide services to big mining operators. Emerging machine learning techniques have a significant potential to improve cost estimation accuracy in mining projects. This thesis will demonstrate the applicability of artificial neural networks (ANN) and support vector regression (SVM) for cost estimation for underground mines. The primary motivation for this research is that the relationship between exploratory and response variables is usually complex and non-linear. Also, no current technique estimates costs accurately. The use of machine learning models for the cost estimation of mining projects is still relatively new, and there is a limited amount of research that has been done in this area.

1.2 Originality

The originality of this work lies in demonstrating the potential of machine learning models, namely, ANN and SVM, to improve the accuracy of cost estimation for mining projects and in showing the feasibility of using these models in a practical setting through case studies. From a contractor perspective, developing more accurate cost estimation models is vital in reducing the risk of cost overruns and improving profitability. Therefore, demonstrating the potential of machine learning models to improve cost estimation accuracy in mining projects can be a valuable contribution to the industry. Underground mining projects will be a particular focus of this research.

1.3 Thesis outline

Chapter 1 defines the problem, objectives, and originality of the thesis.

Chapter 2 conducts a comprehensive literature review. It encompasses an exploration of contractor roles, contract methodologies relevant to mining operations, cost estimation taxonomy, various cost estimation techniques, the structure of bid prices, and an introductory discussion on key machine learning concepts, along with notable applications identified within the existing literature.

Chapter 3 elucidates the methodology employed, delving into crucial aspects of ANN and SVR, such as activation functions, backpropagation, and kernel trick. Also, the selection of hyperparameters and the evaluation metrics are discussed in detail.

Chapter 4, a comprehensive case study involving a mine in The Americas, thoroughly examines, highlighting the application of ANN and SVR. Also, a critical analysis, benchmark, and comparison are provided.

Chapter 5 provides the conclusions drawn from the study and offers recommendations for future research horizon within the scope of this thesis.

2 Literature review

Underground mining is a complex and multi-step process that involves the extraction of valuable minerals or ores from deep below the earth's surface. Underground mining activities include development and production. The development activities refer to creating underground openings or excavations that provide access to mineral deposits. Production activities include drilling, blasting, mucking, and haulage activities. In general, these activities are implemented concurrently in large mines and build on the foundation for the entire mining operation. Depending on deposit depth, dip, orientation, and strike, and mining method, an access network is developed. This network includes shafts, declines, ramps, drifts, cross-cuts, and levels. Also, ventilation, water management, compressor air, surface buildings, and power infrastructure should be installed. This thesis focuses on the development activities, which include a range of unit processes, such as horizontal and vertical access excavations, supporting, caving, and maintenance.

2.1 Contractors

A common business model in the mining industry is to work with contractors. In this model, some elements of mining operations are contracted to third parties called contractors. This thesis focuses on the cases where development activities are given to contractors specifically. The most common services offered in the development of an underground mine are lateral advancement, shaft sinking, raise boring, ramp development, inclines, cross-cuts, and rehabilitation works. These services encompass the essential aspects of mine development, such as creating openings and pathways for access, rehabilitating existing structures, constructing ramps for entry points, developing inclined roadways for lower-level access, excavating horizontal roadways, and carrying out mass excavation activities.

Contracting underground mine development activities to a third party has several benefits:

a. Mine owners may not have sufficient equipment, experience, and critical human accumulation to conduct mine development. Contractors with specialization and equipment fleet for development can provide a business advantage to the mining operator.

b. They can also provide high flexibility to mining operations, allowing companies to adjust their development schedules as needed to meet changing production demands. Also, as mentioned before, underground mining is highly complex and includes many activities. Working with contractors can ensure better control and monitor operations. Thus, the project planning and control will be easier.

c. Contracting is a risk diversification strategy. Some technical risks associated with deviations from production targets and contingencies are transferred to the contractor.

When selecting a contractor for underground mine development, it is important to consider their past safety record. Underground mining is a hazardous activity, and contractors must have a strong safety culture and coincide with the required safety protocols to ensure their workers' safety. It is also essential to consider a contractor's experience and track record in underground mine development, as well as their equipment and resources.

Working with a contractor takes several steps, including tendering, bidding, contract negotiation, and performance. Tendering is the initial phase of the procedure. The mining company Requests for Proposal (RFP) from prospective contractors, inviting them to submit bids to develop/operate the mine. Details about the project's scope, the expected length of the contract, and the performance criteria that the contractor must satisfy are included in RFPs. After reviewing the RFP and creating a bid that contains a thorough plan for how they will carry out the work, the contractors who are interested in developing/operating the mine submit their bids. In addition to the contractor's proposed schedule and budget, the bid will typically include information about the tools and labour they plan to use. The mining company will evaluate the bids and select a qualified contractor to conduct the tasks described in the tendering process. Typically, the selection procedure entails technical and financial evaluation, with the contractor's previous project performance, technical expertise, reputation, tender price submitted, etc. The contractor selection criteria may have differences from industry to industry (Khoso & Yusof, 2020; Olanrewaju et al., 2022; Watt et al., 2010). Contract negotiations follow the selection of a contractor. This step requires the mining company and contractor to come to an agreement on the contract's terms, such as the work's parameters, its cost, and the contractor's expected level of performance. The contract will also detail the terms of payment, the length of the agreement, and penalties for non-performance. The contractor will start performing its duties after the agreement has been signed. All tasks necessary for the mine's operation, such as mine development, production, and maintenance, will fall under the purview of the contractor. Drilling, blasting, supporting, ventilation, scaling, mucking, haulage, and other unit processes are typical duties.

Depending on the specific needs of mine, different equipment and labor will be needed to conduct activities. As an example, a contractor running an underground mine might need specialized equipment (e.g., jumbos, rock drills, bolters, and loaders) and skilled labor (e.g., miners, engineers, technicians, and mechanics). Contracting out offers the advantage of allowing contractor to deliver equipment flexibility as one of its benefits (Awuah-Offei et al., 2003).

Additional support services, such as catering, transportation, and security, might also be required from the contractor.

The contractor shall maintain the performance standards set forth in the contract throughout the term of the contract. These performance requirements will typically include criteria like production goals, safety requirements, and environmental compliance. Penalties or contract termination may be a consequence of failing to adhere to these standards.

In major mining countries, contractors play a significant role. Thanks to their expertise, mining operations are conducted in a more profitable, economical, safer, and timely manner. Also, they provide flexibility to mining corporations. The contractors offer a wide range of services, including exploration, development, operation, mineral processing, and materials handling.

2.2 Contract approaches

The most common contract approaches are cost-reimbursable, lump-sum, and unit-price (Burnett & Wampler, 1998; Ibbs et al., 2003; Picornell et al., 2017).

In a cost-reimbursable contract, the contractor is reimbursed by the owner for all legitimate actual costs incurred, along with a profit addition. In this case, the owner assumes more risks, as they are responsible for paying all the contractor's actual costs (Picornell et al., 2017). Consequently, the owner needs to actively manage and control these costs in line with their planned budget. For the contractor, profit is guaranteed, and it is usually proportional to the costs, fixed, or a combination of both, making it easy to calculate for the contractor.

In lump-sum or fixed-price contracts, the contractor is paid a predetermined price by the owner regardless of the actual expenses incurred. Here, the contractor assumes more risks, and places specific attention on planning and controlling the project effectively. The total project price is fixed and does not deviate from the contractual budget unless modifications are made to the contract (Ibbs et al., 2003). Consequently, the contractor strives to plan the costs as accurately as possible and closely monitor the deviation between planned costs and actual costs. Any additional cost overruns reduce the profitability of the contractor since the contractual price is fixed.

Unit-price contracts involve the owner periodically paying the contractor based on preset unit rates applied to the actual measured quantities. These unit rates include estimated costs, overhead, and profit (Picornell et al., 2017). Unit-price contracts represent a hybrid payment approach that combines features of lump-sum and cost-reimbursable contracts. Risks in unit-price contracts are more balanced between both parties. While quantities may vary during the contract's development based on the actual work, the unit price rates are fixed from the beginning. Both the contractor and the owner bear some risks, which incentivizes both parties to implement effective planning and control procedures. From the owner's perspective, the contractual budget is determined through the contractor's bid and is distributed in periodic payments throughout the project's duration (Burnett & Wampler, 1998). However, unlike a fixed-price agreement, the budget in a unit-price contract can vary based on the measurement of actual quantities.

This thesis is poised to make a substantial contribution to the realm of unit price contracts. The significance of this contribution lies in its potential to enhance the accuracy and reliability of unit price estimations, thereby optimizing cost assessments in various project scenarios.

2.3 Methodologies for cost estimation

The Project Management Body of Knowledge (PMBOK) is a crucial point of reference across various project domains. It encompasses an array of methodologies geared towards the discipline of cost estimation and control. It is important to highlight the inherent uniqueness of each project, necessitating bespoke approaches. Different corporate entities adopt their own distinct cost estimating policies, procedures, or guidelines. PMBOK defines cost estimation as the systematic process of approximating the fiscal outlay required to complete project undertakings. The precision of these estimates typically refines as the project matures. In the preliminary phases, rough order of magnitude estimates exhibits a broad accuracy range, ranging from -25% to +75%, which gradually narrows to a range of -5% to -10% as project-specific information attains a higher degree of granularity. Costs encompass a multifarious spectrum of resources, spanning labor, materials, equipment, services, facilities, and contingencies, among other pertinent elements. Cost estimates can be methodically presented either at the granular activity level or in a more concise, aggregated format.

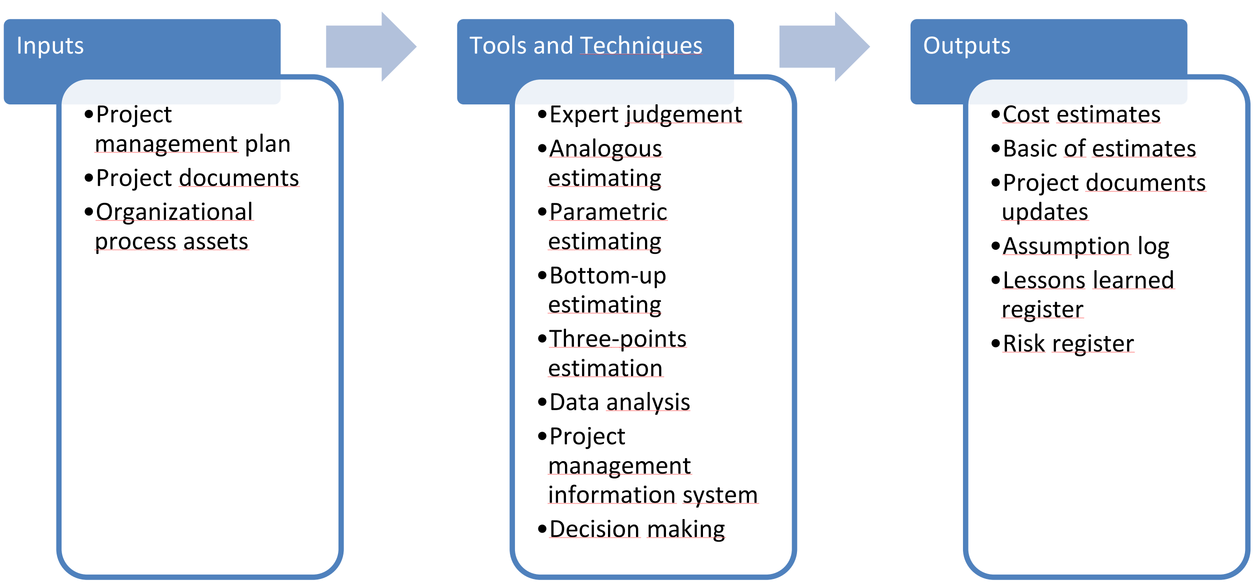

Figure 1 Tools and techniques to estimate costs adapted from (PMI, 2017)

Within the annals of PMBOK, a multitude of methodologies are elucidated in order to calculate the cost estimates as well as their necessary inputs and the outputs provided for their intends, as shown in Figure 1. A project management plan, the organizational process assets, and the project documents, should be given as inputs for the utilization of the techniques. These techniques include:

- Expert Judgment: An approach that calls upon individuals endowed with extensive expertise, informed by experience, specialized education, and a profound understanding of analogous projects and pertinent cost estimation methodologies.
- Analogous Estimating: A method predicated on the empirical analysis of historical data emanating from comparable activities or projects, serving as a benchmark for current estimation exercises.
- Parametric Estimating: A quantitative technique underpinned by algorithms and statistical relationships, employing historical data and project-specific parameters as inputs to ascertain cost estimates.
- Bottom-up Estimating: A detailed estimation methodology that meticulously assesses the cost implications of individual work packages or activities, subsequently consolidating these assessments for higher-level reporting and monitoring purposes, often hinging on the Work Breakdown Structure (WBS).
- Three-point Estimating: A technique that endeavors to enhance the accuracy of single-point cost estimations by considering the spectrum of estimation uncertainty and risk, relying on three distinct estimates to define a comprehensive range for the projected cost of a given activity. This triumvirate of estimates encompasses the most likely, optimistic, and pessimistic scenarios.

- Data Analysis: This encompasses various analytical techniques, notably alternatives analysis and reserve analysis, which are leveraged to assess multiple options and factors such as cost uncertainties.
- Cost of Quality: This dimension extends to considering the implications of quality-related expenditures versus non-conformance costs, encompassing both short-term cost reductions and the potential long-term ramifications during the product life cycle.
- Project Management Information System (PMIS): A repository of software tools that expedite cost estimation and analysis, streamlining the application of cost-estimation techniques and facilitating the expeditious evaluation of cost estimate alternatives.
- Decision Making: Within the sphere of decision-making, the most frequently employed technique is voting, which involves the evaluation of multiple alternatives, with expected outcomes manifesting in the form of future actions. These methodologies play a pivotal role in harnessing team input to augment the accuracy of estimates and enhance team commitment.

2.4 Types of estimates

Cost estimation involves a meticulous evaluation of all the costs associated with the various elements of a project or effort, all defined within an agreed-upon scope. This financial assessment is essential for effective project planning and execution. The American National Standards Institute (ANSI) has delineated three specific types of cost estimation, categorized based on the degree of definition within the process industry:

- Order-of-magnitude estimates: This category of estimates is characterized by an expected accuracy range of between +50% and -30%. Typically, order-of-magnitude estimates are grounded in cost-capacity curves and cost-capacity ratios. What sets them apart is their independence from preliminary design work. These estimates provide an initial understanding of the financial landscape of a project, serving as a valuable starting point.

- Budget estimates: In contrast, budget estimates delve deeper into the intricacies of project costs. They draw from tangible data such as flowsheets, layouts, preliminary equipment descriptions, and specifications. The expected accuracy of budget estimates falls within a range of +30% to -15%. To perform accurate budget estimates, it is typically necessary for the project design to have progressed to a 5% to 20% completion stage. This level of detail is vital for more advanced financial planning and decision-making.
- Definitive estimates: Definitive estimates represent the highest degree of precision in cost estimation. They necessitate the acquisition and integration of meticulously defined engineering data, including site-specific information, detailed specifications, basic drawings, intricate sketches, and accurate equipment quotations. These estimates are typically prepared when the project design has reached a level of completeness ranging from 20% to a full 100%. The hallmark of definitive estimates is the rigorous accuracy, falling within the range of +15% to -5%. They provide a rock-solid foundation for advanced financial planning and informed decision-making, especially in complex projects.

The field of cost estimation has evolved to adapt to the intricacies of modern projects. Association for the Advancement of Cost Engineering (AACE), a prominent authority in cost engineering, has proposed an extension of ANSI's estimate classifications, introducing five distinct types. Each type aligns with the level of project definition at the time of estimation and comes with an expected accuracy level.

Accuracy in each estimate class depends on factors such as the technological complexity of the project, the availability of relevant reference information, and the inclusion of appropriate contingency provisions. It is worth noting that, in exceptional circumstances, accuracy levels may deviate from the conventional ranges associated with each estimate class.

In addition to their work on estimating classifications, AACE International has developed a suite of Recommended Practices (RPs) tailored to diverse industries. Notably, within the Mining and Mineral Processing sector, AACE International has formulated RP 47R-11. This resource offers detailed descriptions of the five estimate classes:

- Class 5 estimate: Typically rely on inferred resources, established through preliminary drilling exploration and metallurgical analysis, along with insights gained from related projects to meet disclosure standards like NI 43-101. These estimates encompass a basic geological model, mineralization grade estimates, and an initial mine plan. This plan includes essential details such as mining methods (e.g., open pit or underground), gross production schedules, nominal plant capacity, assumed block flow diagrams, process rates, and a conceptual infrastructure outline. While there may be minimal metallurgical test work and auxiliary studies available, no detailed design drawings or equipment specifications are typically prepared, beyond preliminary notes and sketches by the project engineer, which may include little more than the proposed plant type, capacity, and location.
- Class 4 estimate: Is usually based on indicated or measured resources confirmed through detailed drilling. These estimates necessitate the development of a

preliminary geological model and a comprehensive mine plan, encompassing elements such as pit optimization, geotechnical assessments, hydro-geological studies, and more. Metallurgical test work aims to determine the probable process flow sheet, approximate material balance, and the identification of major equipment. Engineering efforts comprise, at a minimum, the creation of general arrangement drawings, equipment lists for major components, specifications for nominal plant capacity, block schematics, and process flow diagrams for the main process systems.

- Class 3 estimate: Is prepared using probable or proven ore reserves, adhering to defined confidence limits as specified by securities codes. A detailed mine plan is a prerequisite, with the potential for pre-stripping activities to commence upon project approval. Metallurgical test work reaches a level of detail that allows for expansion in equipment lists and specifications. Engineering is expected to provide general arrangement drawings, preliminary piping and instrument diagrams, single-line electrical drawings, and more refined plot plans and layout drawings.
- Class 2 estimate: Is based on a detailed mine design and production schedule, leveraging proven ore reserves. These estimates encompass all information included in Class 4 and Class 3 estimates concerning process details. Furthermore, as the project progresses toward the execution phase, Class 2 estimates include additional engineering data, such as a comprehensive project execution plan, detailed schedules, precise topographic maps, on-site surveys to ascertain plant site parameters, and foundational data.

- Class 1 estimate: Class 1 estimates represent the highest level of detail in estimation and are typically prepared for discrete project segments, rather than the entire project. These estimates rely on proven reserves, often with civil and pre-stripping activities already underway. Class 1 estimates provide extensive detail and may serve as references for subcontractor bidding or for various assessments and decisions made by project owners.

Table 1 provides a detailed description of these estimate classifications.

Table 1 Cost Estimate Classification Matrix for Mining and Mineral Processing Industries taken from (AACE, 2020)

	Primary Characteristic	***Secondary Characteristic***		
ESTIMATE CLASS	**MATURITY LEVEL OF PROJECT DEFINITION DELIVERABLES** Expressed as % of complete definition	**END USAGE** Typical purpose of estimate	**METHODOLOGY** Typical estimating method	**EXPECTED ACCURACY RANGE** Typical variation in low and high ranges at an 80% confidence interval
Class 5	0% to 2%	Conceptual planning	Capacity factored, parametric models, judgement, or analogy	L: -20% to -50% H: +30% to +100%
Class 4	1% to 15%	Screening options	Equipment factored or parametric models	L: -15% to -30% H: +20% to +50%
Class 3	10% to 40%	Funding authorization	Semi-detailed unit costs with assembly level line items	L: -10% to -20% H: +10% to +30%
Class 2	30% to 75%	Project control	Detailed unit cost with forced detailed take-off	L: -5% to -15% H: +5% to +20%
Class 1	65% to 100%	Fixed price bid check estimate	Detailed unit cost with detailed take-off	L: -3% to -10% H: +3% to +15%

Must be highlighted that the maturity level of definition is the sole determining (i.e., primary) characteristic of class. In the Table 2, the maturity level of the project definition deliverables can be defined, considering that a list of the documents and its status for each class of estimate, is provided. So, can be observed for example that the Mine Operations Layout it is in its preliminary status for a class 4 or class 3 estimation, but for a class 2 estimation should be defined.

The maturity level is defined as an approximation of the completion status of the respective deliverables. The completion status is indicated by the following letters (AACE, 2020).

• None (Blank): Development of the deliverable has not begun.

• Started (S): Work on the deliverable has begun. Development is typically limited to sketches, rough outlines, or similar levels of early completion.

• Preliminary (P): Work on the deliverable is advanced. Interim, cross-functional reviews have usually been conducted. Development may be near completion except for final reviews and approvals.

• Complete (C): The deliverable has been reviewed and approved as appropriate.

Table 2 Maturity level of project definition deliverables taken from (AACE, 2020)

	CLASS 5	**CLASS 4**	**CLASS 3**	**CLASS 2**	**CLASS 1**
MATURITY LEVEL OF PROJECT DEFINITION DELIVERABLES	**0% to 2%**	**1% to 15%**	**10% to 40%**	**30% to 75%**	**65% to 100%**
General Project Data:					
Project Scope Description	General	Preliminary	Defined	Defined	Defined
Mine and Plant Production/Facility Capacity	Assumed	Preliminary	Defined	Defined	Defined
Plant Location	General	Approximate	Specific	Specific	Specific
Soils & Hydrology	None	Preliminary	Defined	Defined	Defined
Resource Determination	Inferred	Indicated	Measured	Measured	Measured
Reserve Determination	Assumed	Probable	Proven	Proven	Proven
Geology	General	Preliminary	Defined	Defined	Defined
Geotechnical and Rock Mechanics	General	Preliminary	Defined	Defined	Defined
Metallurgical Teste Work	General	Preliminary	Defined	Defined	Defined
Integrated Project Plan	None	Preliminary	Defined	Defined	Defined
Project Master Schedule	Assumed	Preliminary	Defined	Defined	Defined
Mine Life Plan/Schedule	General	Preliminary	Preliminary	Defined	Defined

Initial Mine/Ore Access (Roads, Prestripping, Tunnels, Shafts, Water Management, Waste Management, etc.)	General	Preliminary	Defined	Defined	Defined
Mine Operations Layout (Pit Design, Dumps, Roads, Water Management, Waste Management, etc.)	General	Preliminary	Preliminary	Defined	Defined
Escalation Strategy	None	Preliminary	Defined	Defined	Defined
Work Breakdown Structure	None	Preliminary	Defined	Defined	Defined
Project Code of Accounts	None	Preliminary	Defined	Defined	Defined
Contracting Strategy	Assumed	Assumed	Preliminary	Defined	Defined
Mine (Production Equipment, Prestripping, etc.)	Assumed	Preliminary	Defined	Defined	Defined
Non-Process Facilities (Infrastructure, Ports, Pipeline, Power Transmission, etc.)	Assumed	Preliminary	Defined	Defined	Defined
Engineering Deliverables:					
Block Flow Diagrams	S/P	P/C	C	C	C
Plot Plans		S/P	P	C	C
Process Flow Diagrams (PFDs)		P	C	C	C
Utility Flow Diagrams (UFDs)		S/P	C	C	C
Piping & Instrument Diagrams (P&IDs)		S/P	C	C	C
Heat & Material Balances		S/P	C	C	C
Process Equipment List		S/P	C	C	C
Utility Equipment List		S/P	C	C	C
Electrical One-Line Drawings		S/P	C	C	C
Specifications & Datasheets		S	P/C	C	C
General Equipment Arrangement Drawings		S	C	C	C
Spare Parts Listings			P	C	C
Mechanical Discipline Drawings			S/P	P/C	C
Electrical Discipline Drawings			S/P	P/C	C
Instrumentation/Control System Discipline Drawings			S/P	P/C	C
Civil/Structural/Architectural Discipline Drawings			S/P	P/C	C

During the unit price analysis, costs are estimated for each resource used, incorporating various methodologies summarized in Figure 2.

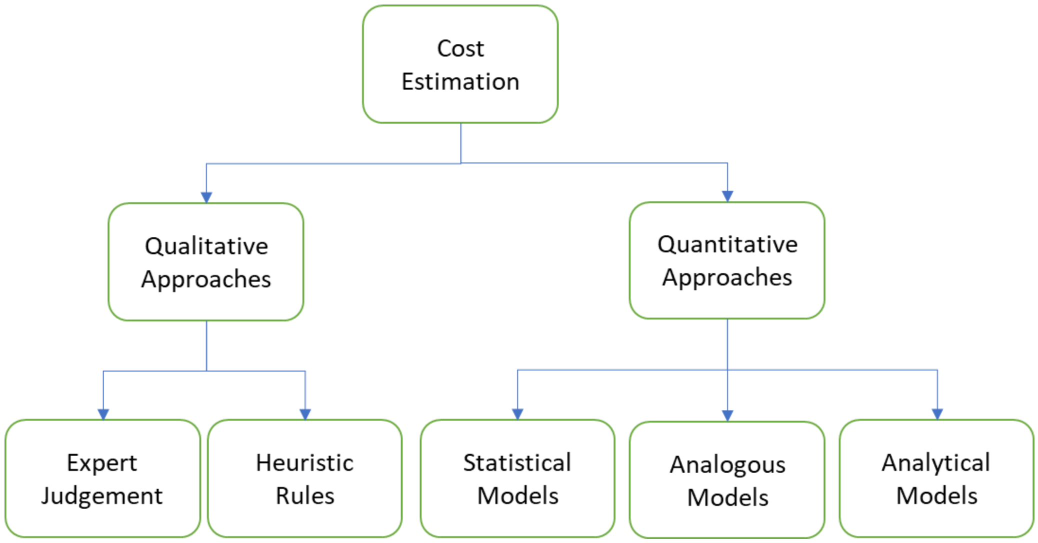

Figure 2 Cost estimation methodologies, adapted from (Caputo & Pelagagge, 2008; Layer et al., 2002)

The literature provides a variety of methods for cost estimation, which can be classified into qualitative and quantitative approaches (Layer et al., 2002). While qualitative approaches relying on expert judgment consider expertise from a group or person with training, specialized education, knowledge, skill or experience in previous similar projects (PMI, 2017), quantitative approaches make use of mathematical models to predict costs. As seen in Figure 2, the quantitative methods can be classified into analytical, analogous, and statistical models (Caputo & Pelagagge, 2008), each with pros and cons. However, the accuracy of these models is not always satisfactory, and there is a risk of financial losses for a contractor.

Regards to the qualitative methods, expert judgment entails harnessing the wealth of knowledge and experience possessed by seasoned professionals in the field. This approach involves seeking insights and opinions from individuals with a deep understanding of the project or domain. While it provides valuable qualitative inputs, its subjective nature may introduce biases, and the reliability of the judgment is contingent upon the expertise of the individuals involved.

In contrast, heuristic rules offer a simplified decision-making framework based on past experiences and generalizations. These rules serve as practical, easy-to-follow guidelines that enable quick cost estimates without the need for complex calculations. Heuristic rules are particularly applicable in situations where historical data is reliable, although their reliance on simplifications may lead to inaccuracies, especially in complex scenarios. There are some rules of thumb well known in the mining field proposed by O’Hara in the 90s (T. Alan O'Hara, 1992), but as these rules were developed with information that have not been updated since then, is not recommended to use them.

This thesis endeavors to contribute to the enhancement of statistical models commonly employed in cost estimation practices. While traditional approaches often rely on linear regression techniques, this study puts forth a pioneering exploration into the application of advanced methodologies, specifically ANN and SVM. The motivation behind this deviation from conventional practices stems from the recognition that the intricacies of cost estimation warrant a more nuanced and adaptive modeling approach.

2.5 Bid price structure.

Figure 3 provides a breakdown of the components of project cost from the contractor's perspective. The project cost encompasses both indirect and direct costs.

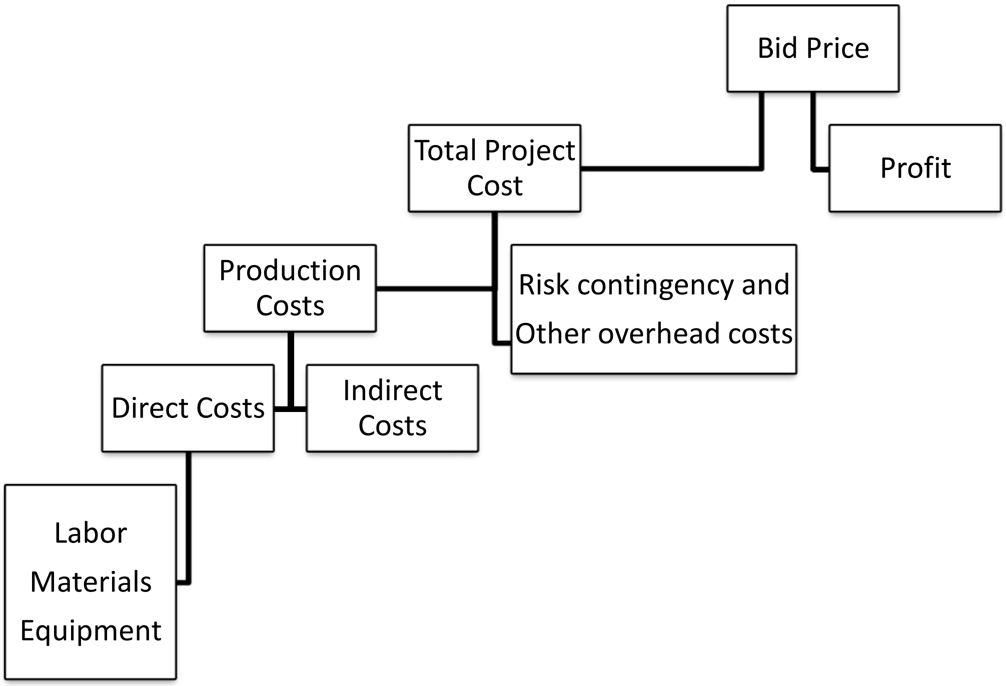

Figure 3 Typical bid price structure in projects adapted from (Sanaz Tayefeh et al., 2020)

The direct cost of the project consists of expenses directly incurred in project and production activities. Typically include labor costs as drillers, helpers, and operators. Also include equipment, materials, and subcontractors who work on specific work packages under the general contractor’s purview.

Indirect costs can be further classified into some categories: project overhead and general overhead (Sanaz Tayefeh et al., 2020). Project overhead costs are indirectly incurred expenses associated with the project that cannot be directly assigned to specific work packages. Examples include utilities, supervision, additional support such as planning, safety, human resources, and similar costs. On the other hand, general overhead costs are not attributable to individual projects but rather pertain to the overall expenses incurred by the contractor's organization. These costs typically encompass expenditures incurred at the head office, personnel costs, and other similar expenses. General overhead costs are allocated proportionately to projects based on their respective contributions to the contractor's total costs.

The contractor's bid price also includes the project's cost and the desired profit. The specific value of profit depends on numerous factors, such as business objectives, competition within the industry, and the contractor's determination to secure the project over its rivals.

Risk contingency is an allocated amount intended to cover risks that may arise during project execution. It serves as an estimated reserve to oversee unforeseen circumstances.

A contractor managing specific tasks will typically base their business strategy on several variables, including the contract price, labor costs, and mining performance. While achieving the performance standards outlined in the contract, the contractor will work to minimize costs and maximize profit. This could require putting into practice budget-friendly strategies like streamlining procedures, reducing costs through organization and management, and improving equipment utilization.

Contractors operate mines more effectively and efficiently than mining companies that might not have the same level of expertise (Suglo, 2010). Additionally, contractors give mining operations a degree of adaptability by enabling businesses to change their development schedules to meet shifting production demands. Furthermore, contracting out a mining operation can aid mining companies in cost management and lower capital expenditures.

This research aims to develop a mine development cost estimation model for mining contractors. Cost estimation is a critical part of any mining project, including the development and operation of underground mines. It involves estimating costs associated with the stages of the mining process, from exploration to production, and accounting for factors such as equipment, labor, materials, and transportation.

The cost estimation process typically begins with developing a mining plan, which outlines the key steps involved in the project and the resources required to carry them out. The plan is then used as the basis for estimating the costs associated with each step, and these costs are aggregated to produce an overall project cost estimate.

In a unit price contract, for example, when a contractor is responsible for mine development, the contract criteria typically are based on the advanced linear meter, a cubic meter of excavation, or tonnage moved. To perform the unit price analysis, as in any project, the more information available, the more accurate the estimation will be. One of the crucial elements during an assessment is to analyze a clear understanding of the project scope. With the project scope and key mine activities, important aspects are clarified, including the project duration, expected progress rates, the types of faces to be advanced, their dimensions, rock types, and support requirements, which are essential knowledge for this analysis. With this information, the contractor can determine the necessary resources for project execution, including the equipment, personnel, and materials to be used, and conduct an estimation process by face and rock types.

When analyzing direct costs, it is important to categorize equipment, materials, and labor resources correctly. Once all the resources are identified, they need to be converted to cost per unit of sale. For example, a contractor is responsible for vertical development, where the advanced linear meter is typically the unit of sale. The resources needed for the lineal advance in an underground mine to calculate the direct cost are listed in Table 3.

Table 3 Resource description example to estimate direct costs.

EQUIPMENT	MATERIALS	LABOR
Description	**Description**	**Role**
Major equipment	**Drilling steel**	Operations Supervisor
Jumbo	Drill rod	Jumbo operator
Bolter	Drill bits	Jumbo operator helper
Scoop	Reaming drill bit	Scoop operator
Telehandler	Pilot adaptors	Dumper operator
Dumper	Shank Adapter	Drilling machine operator
Minor equipment	Coupling	Drilling machine helper
ANFO loader	**Services**	Shotcrete operator
Pneumatic pump	Tools	Shotcrete helper
Jackleg driller	Pipes HDP (water/air/pumping)	Mixer operator
Drill bit sharpener	Pipe fittings	Telehandler operator
	Hanger Hook	Mechanic
	Ventilation ducts	Electromechanic
	Accessories for ventilation ducts	Welder
	Support	Tire man
	Rockbolt	
	Electro-welded mesh	
	Split set	
	Shotcrete	

For equipment, conducting a cycle analysis is crucial to obtain the expected performance of each piece of equipment per advanced meter. Additionally, the equipment's hourly cost must be considered, which will differ for the contractor depending on whether they own the equipment or rent it from a third party. Based on this information, the cost per meter for all equipment can be calculated.

For materials, it is necessary to have the support and service standards for each face, as they will identify the materials to be used during the advance and calculate the required units per meter. The cost of each element is essential information. For drilling steels, it is vital to calculate how much steel will be used, i.e., how many meters will be drilled per advanced meter, and have a reliable service life provided by the steel distributor. With this information, the calculation per meter for each element can be determined, resulting in a subtotal of the material cost.

For labor costs, it is useful to have cycle analysis to determine the number of hours dedicated by role in each face and consider the number of faces where the same person will be employed during their shift, as it will significantly impact the labor cost calculation. Additionally, the costs associated with each position must be considered, as it includes wages, benefits, and cuts. Thus, the labor cost per meter can be calculated.

Within the analysis of indirect costs, the technical and administrative personnel in the project (e.g., management, operation supervision, planning and cost, safety, human resources, maintenance, etc.) are considered, as well as aspects such as system and communication equipment, support vehicles such as vans or buses, accommodation and associated services like cleaning, laundry, and meals. Infrastructure costs include constructing buildings where personnel will work within the project and operation is maintained. Safety implements such as personal protective equipment, signage, and fire extinguishers. It is also important to consider financial expenses such as surety bonds and liability insurance.

2.6 Machine learning in cost estimation

Accurate cost estimation is essential for effective decision-making, including determining the viability of a project and optimizing resource allocation (Gould, 2018). However, traditional cost estimation methods often lack accuracy due to their reliance on limited historical data (Meng and Whittle, 2020). Machine learning techniques are promising in enhancing the accuracy and efficiency of cost estimation in various industries, including mining.

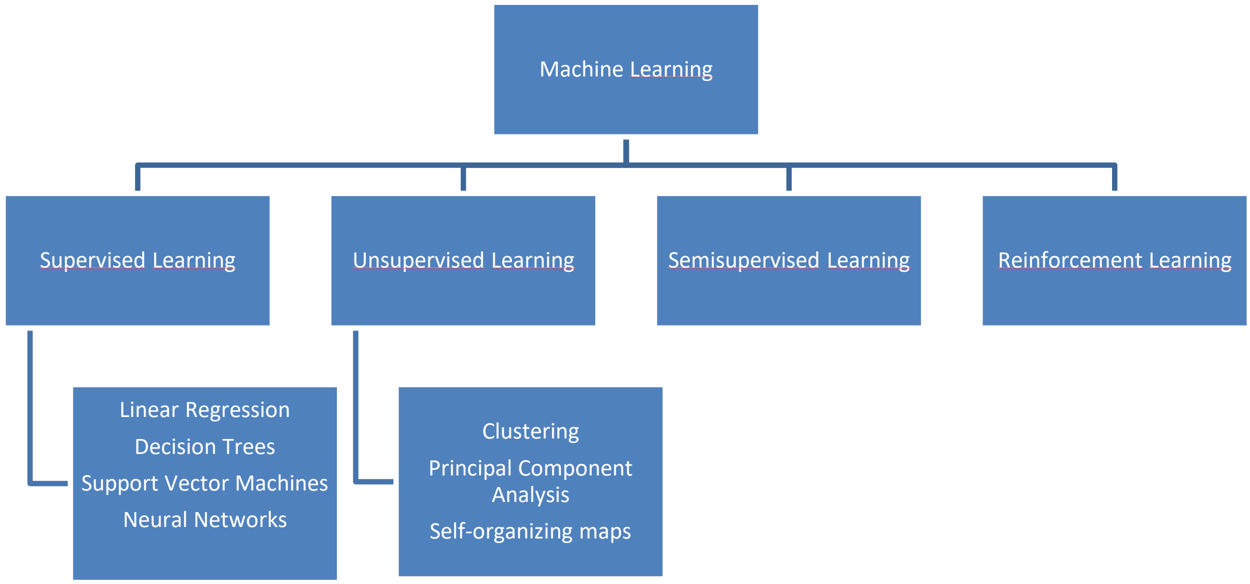

Figure 4 Machine Learning categories

Machine learning is a subfield of artificial intelligence that focuses on the development of algorithms and models that enable computers to learn from and make predictions or decisions based on data (Bishop & Nasrabadi, 2006). At its core, machine learning is about creating systems that can automatically improve their performance on a specific task through the process of learning from data, trying to imitate the way humans learn, improving accuracy gradually. The term "Machine learning" was first introduced by Arthur Samuel with his research where he investigated two machine learning procedures in the game of checkers, and it was the first time a computer playing a game won against a human ever (Samuel, 1959).

There are some categories where we can divide the machine learning techniques, as is shown in Figure 4. The main categories are supervised and unsupervised learning. The main characteristic of supervised learning is the availability of labeled examples in the training data (Cunningham et al., 2008), so that in the test set the unlabeled examples can be identified with the highest possible accuracy. It is commonly used for classification and regression analysis. In the other hand, the unsupervised learning has not access to labeled examples, and is commonly used in clustering problems (Liu, 2011).

There are another methods like the semi-supervised learning, where both labeled and unlabeled examples are used to generate the prediction (Nasteski, 2017), and the reinforcement learning, which deals with introducing foundational classes of algorithms for learning optimal behaviors, based on various definitions of optimality with respect to the goal of learning sequential decisions (Wiering & Van Otterlo, 2012).

In the literature, there is a limited availability of articles discussing the application of machine learning to cost estimation in mining. A few examples found include the following:

Hamidreza and Morteza (Hamidreza & Morteza, 2019), applied support vector regression (SVR) to determine the mining capital expenditure. The variables they recommended to estimate were annual mine production, stripping ratio, the annual production of the mill, reserve mean grade given in % Cu, and the mine life. They used the data from 52 open-pit copper mines and compared with the kernel ridge regression. Results showed that both methods had good results in the training stage, but the support vector regression was far better than the kernel ridge regression in the testing stage, showing an acceptable range of error.

The other machine learning models were applied to predict mining capital cost, applying ANN approach (Hongquan et al., 2019). The authors used a dataset of 74 observations obtained from 74 open-pit copper mining projects, and the variables used to make the estimation were those proposed by Hamidreza and Morteza (2019). They trained some machine learning models, including ANN with three different architectures, random forest, SVM and classification and regression tree, and found that the technique that had better results was ANN in all its architectures, followed by the SVM model.

Machine learning has also been applied to cost estimation in other areas as construction, where the application of these algorithms has increased with time (Ran et al., 2022), showing a preference to apply ANN in these kinds of projects (Sanaz Tayefeh et al., 2020). Even with this increase in the application of machine learning algorithms, deep learning algorithms do not attract many researchers. Nevertheless, these studies have demonstrated the potential of machine learning in enhancing the accuracy and efficiency of cost estimation for construction projects.

The mine cost estimating process lacks a widely accepted and rigorous approach, unlike the process employed in industries such as building construction. In mining, this process demonstrates noticeable variations from one evaluation to another, encompassing differences in both approach and scope (SME, 2011). Nonetheless, as mentioned in the previous chapters, there exist international codes and standards that outline the phases of a mining project, including specific types of studies. The internationally recommended practice established by the AACE (AACE, 2020) aims to establish a connection between these study types and the five standard estimate classes utilized in other industries. For instance, Class 5 represents projects with a project definition deliverable maturity ranging from 0% to 2%, while Class 1 corresponds to a maturity level between 65% and 100%. Accuracy expectations within the industry vary for each class, with the accuracy range increasing as the class and the amount of project information obtained expand. A typical accuracy range for a Class 1 estimate, which serves as the final estimate, falls between -10% and 15%. The mining industry experiences considerable time gaps between these phases, which are utilized to gather more information and enhance estimation accuracy. However, for contractors, the situation differs as they are required to provide estimates within months and present a final estimate for tendering purposes. Consequently, it is crucial to focus on refining the methodologies employed in cost estimation, reducing errors within industry-accepted ranges, and thereby avoiding both overestimation and underestimation to mitigate potential losses.

It is difficult to say that a cost estimation approach is superior to the others. Each project has its dynamics that a specific approach makes favorable. These dynamics can be related to data availability, knowledge and experience accumulation, and project characteristics. Mining operations have distinct properties. Firstly, each mine is located in a different geological setting, with varying rock formations, ore deposits, and mineral grades. This uniqueness affects the mining process, equipment required, and, consequently, the cost of mining. Additionally, the regulatory and environmental requirements of each mine differ from one another, making each mine's cost estimation process unique. These factors combine to make each mine project unique, with varying risks and rewards.

From a contractor's perspective, the unique nature of each mine poses a significant risk, as it requires specific equipment, processes, and expertise, which may not be readily available. Consequently, contractors may need to invest more time and money to understand the mine's unique characteristics and requirements, increasing the project's cost. Contractors are sensitive to cost, as higher costs can negatively impact their profits and competitiveness in the marketplace.

3 Methodology

This chapter serves as the blueprint for the practical application of advanced techniques in cost estimation, specifically, ANN and SVR. As the foundation for this research, this section outlines the systematic approach we will follow to leverage the capabilities of ANN and SVR in addressing the complex and critical task of cost estimation within the mining industry.

We will start by explaining how we get our data ready for analysis. Then, we will show you how we create and fine-tune the ANN and SVR models. Finally, we will discuss how we measure the models' accuracy and reliability.

Additionally, the methodology will elucidate the evaluation criteria and techniques employed to gauge the accuracy and effectiveness of our models. This will involve the use of appropriate performance metrics and validation methods to assess the models' predictive capabilities and robustness.

3.1 Artificial Neural Networks (ANN)

ANN has shown tremendous growth in recent years and has been used to solve challenging problems such as object detection, semantic segmentation, person re-identification, image retrieval, anomaly detection, skin disease diagnosis, etc. To learn abstract features from data, diverse types of neural networks have been identified in deep learning, including Multilayer Perceptron, Convolutional Neural Networks, Recurrent Neural Networks, and Generative Adversarial Networks. Some key aspects of neural networks for their effective functioning include weight initialization, loss functions, different layers, overfitting, and optimization (Abu-Doush et al., 2023; Dubey et al., 2022).

ANN, in general, is a computational model inspired by the human brain's structure and function. Multilayer Perceptron (MLP) is a type of feedforward neural network (FNN) in which the data is transferred in one direction only, from the input layer to the output layer (Abu-Doush et al., 2023). They form the layers of interconnected nodes (also called neurons) that receive input signals, process them, and generate output signals (Christopher, 2006). As observed in Figure 6, the nodes in each layer are connected to nodes in the next layer, and a set of weights determines the strength of the connections between nodes. The neurons located in the hidden layers construct linear combinations (z_j) of the input variables, and each one is transformed using a differentiable and non-linear activation function $f(z_j)$.

$$z_j = \sum_{i=0}^{n} x_i . w_{i,j} + b \tag{1}$$

$$a_j = f(z_j) \tag{2}$$

where x_i is the i^{th} feature used as input in the network, $w_{i,j}$ is the weight linked from the i^{th} input to the j^{th} hidden neuron, b is the bias associated with the hidden neuron, z_j is the linear combination done in each neuron, and a_j is the activation function selected to add the non-linearity to the process. A scheme of how each neuron is working is shown in Figure 5.

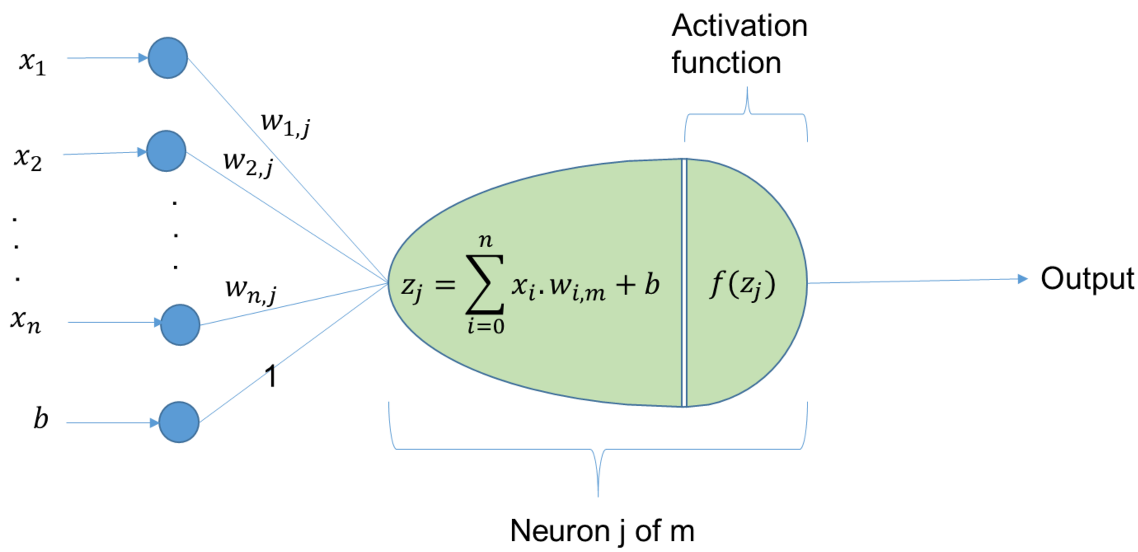

Figure 5 Working scheme of a neuron.

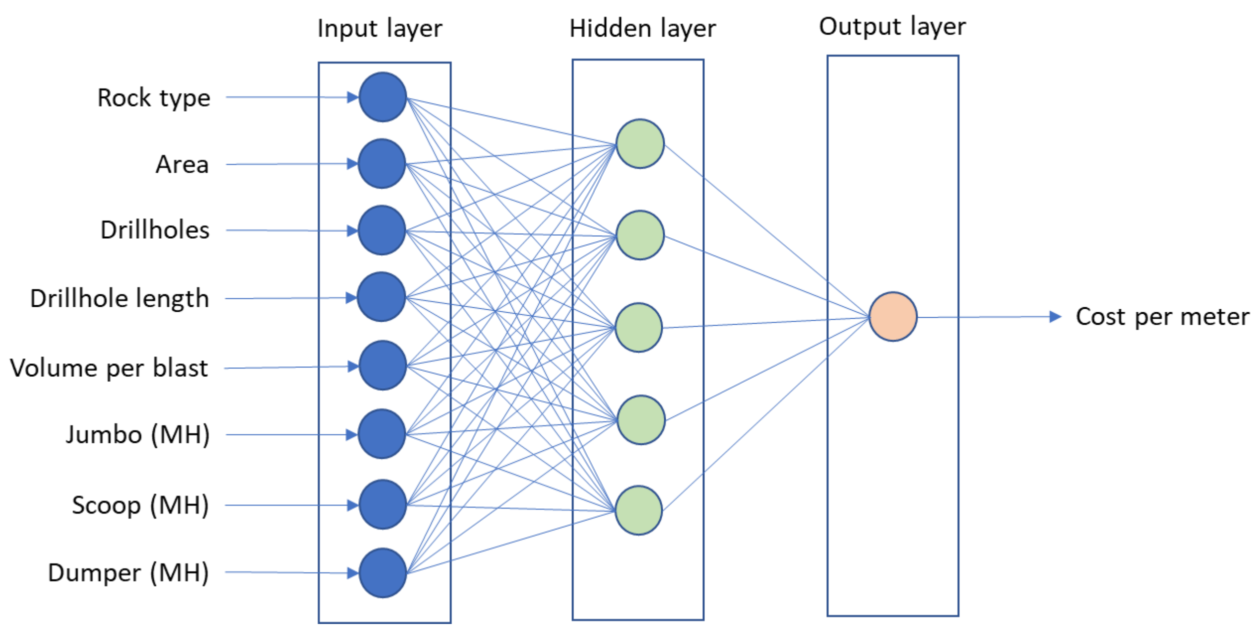

Figure 6 General architecture for ANN for estimating the cost per meter in this study.

3.1.1 Activation Functions

Activation functions are essential in neural networks as they facilitate learning abstract features by applying non-linear transformations to the input data (Dubey et al., 2022). In implementing neural networks, multiple activation functions have been developed to reduce the loss in the training process. A recent study showed at least 25 activation functions that can be divided into seven families: ReLU, Sigmoid, Tanh, Softmax, GELU, Softsign, and Shrink (Singh et al., 2023). Sigmoid and hyperbolic tangent functions have been the most used activation functions in ANN for years (Apicella et al., 2021). Although these bounded functions usually have values between 0 and 1 or -1 and 1 as shown in Figure 7 and perform well in small neural network architectures, but they may suffer from the vanishing gradient problem, which is a problem generated during the training phase when the gradient is used to update the weights of the network. The network cannot learn effectively if the gradients are too small or zero, resulting in poor predictive performance. On the other hand, networks utilizing unbounded functions as the rectified family, such as ReLU, Leaky ReLU, etc., have shown to be universal approximators, effectively mitigating the vanishing gradient problem, because it is only saturated in one direction as can be observed in Figure 8 (Apicella et al., 2021). In Table 4, the most common activation functions are mentioned, with their derivatives and ranges.

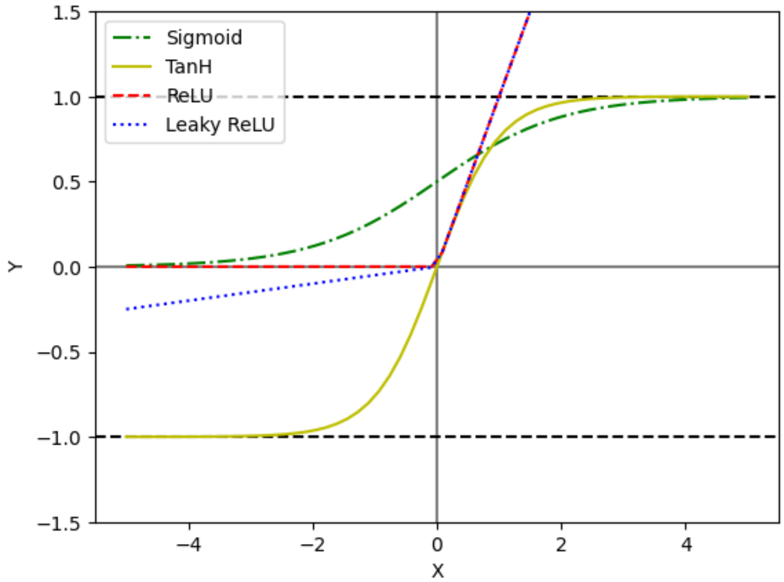

Figure 7 Common activation functions

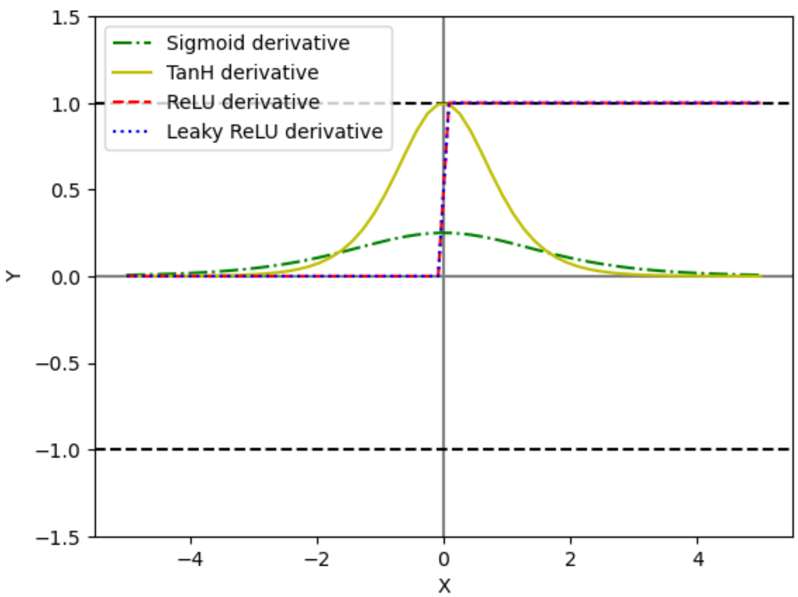

Figure 8 Derivatives of the common activation functions

Table 4 Common activation functions and their derivatives

Function	Definition	Derivative	Range
Sigmoid	$\sigma(x) = \frac{1}{1+\exp\,(-x)}$	$\frac{1}{1+\exp(-z_j)}.\left(1-\frac{1}{1+\exp(-x)}\right)$ $=\ x(1-x),$	[0, 1]
Hyperbolic tangent	$TanH(x)$ $= \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$	$1 - TanH^2(x)$	[-1, 1]
Rectified linear unit	$\mathrm{ReLU(x)} = \max\,(x, 0)$	$1\ if\ x \geq 0$ $0\ if\ x < 0$	[0, ∞]
Leaky ReLU	$\max(\alpha\, x, 0)$	$1\ if\ x \geq 0$ $\alpha\ if\ x < 0$	[-∞, ∞]

3.1.2 Backpropagation

To train an ANN, a dataset is provided to the network with inputs and the corresponding desired outputs. The network processes the inputs and produces an output. The error between the output produced by the network and the desired output is then calculated (Yoosefzadeh Najafabadi et al., 2023). The weights of the connections between nodes are then adjusted to minimize the error, a process known as backpropagation, which backpropagates the error according to the chain rule (Tao et al., 2022).

Backpropagation is a popular algorithm used in ANN for supervised learning. The algorithm enables the network to learn from a given set of input-output pairs and adjust its parameters to minimize the difference between the predicted and actual output values. The backpropagation process involves two stages: forward propagation and backward propagation.

During the forward propagation stage, the input data are passed through the network, and the output of each layer is calculated using the weights of the connections between the neurons. The output of the final layer is then compared to the desired output, and the error is calculated using a chosen loss function.

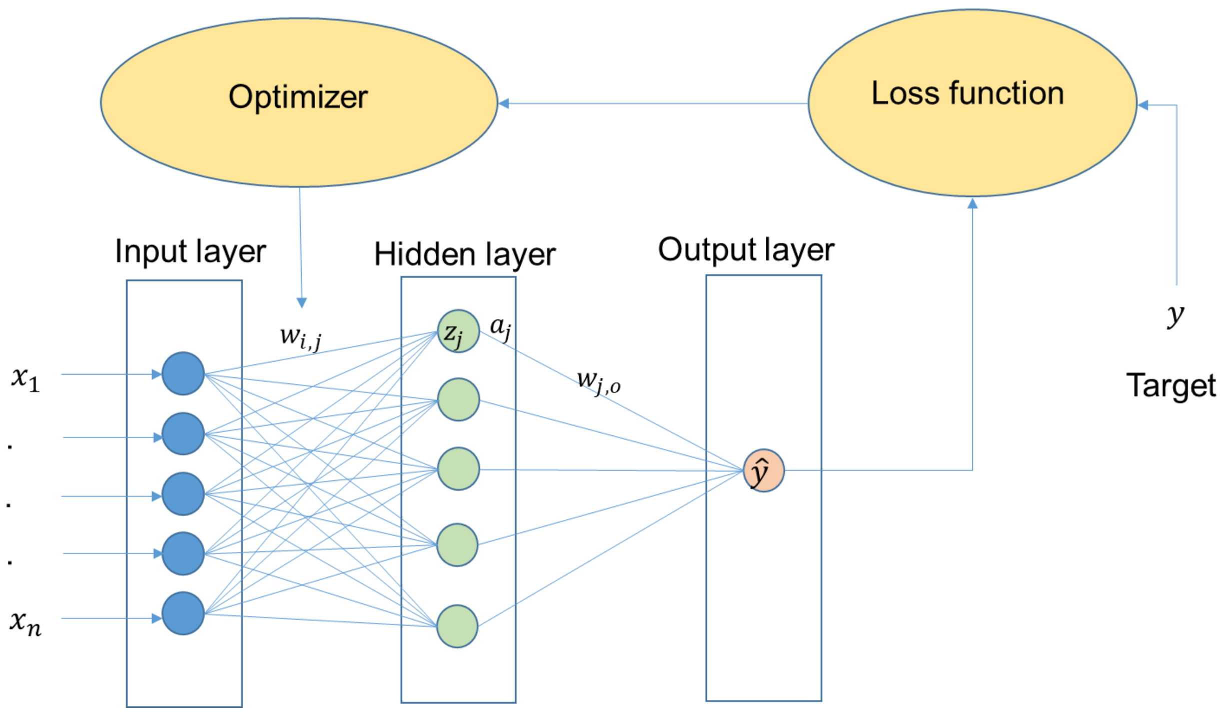

Figure 9 Working scheme of ANN with a single hidden layer.

The output of the final layer is calculated as follows:

$$\hat{y} = \sum_{j=0}^{m} a_j . w_{j,o} + c \tag{3}$$

where $\hat{y}$ is the predicted value, $w_{j,o}$ is the weight linked from the j^{th} hidden neuron to the output, m is the number of neurons and c is the bias associated with the output.

Therefore, the forward pass the ANN is feed by equations (1), (2), (3) and a Loss function choosen during the implementation process, the involves the target and predicted values comparisons.

In the backward propagation stage, the error values are propagated back through the network, starting from the final layer, and moving backwards. The gradient of the error with regard to the weights of the network is calculated for each layer, using the chain rule of calculus (Tao et al., 2022). This involves calculating the partial derivatives of the output of each neuron with respect to the output of the previous layer and the partial derivatives of the error concerning the output of each neuron in the last layer. Therefore, using the chain rule, the partial derivative of the loss function according to the weights of the input given by the following expression is obtained:

$$\frac{\partial L}{\partial w_{i,j}} = \sum \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a_j} \frac{\partial a_j}{\partial w_{i,j}} \tag{4}$$

As can be observed in Figure 9, the activation function involves the z_j function, and for the calculation of each z_j, all the weights $w_{i,j}$, we sum over all of them, having as a result the following equation:

$$\frac{\partial L}{\partial w_{i,j}} = \sum_{n} \sum_{o} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a_j} \frac{\partial a_j}{\partial z_j} \frac{\partial z_j}{\partial w_{i,j}} \tag{5}$$

Let’s say in a process a sigmoid activation function is chosen and the mean square error as loss function, which means we have equations (1), (3) and

$$a_j = f(z_j) = \sigma(z_j) = \frac{1}{1 + \exp{(-z_j)}} \tag{6}$$

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2 \quad (7)$$

Calculating all these partial derivatives:

$$\frac{\partial z_j}{\partial w_{i,j}} = x_i \quad (8)$$

$$\frac{\partial a_j}{\partial z_j} = \frac{1}{1 + \exp(-z_j)}.\left(1 - \frac{1}{1 + \exp(-z_j)}\right) = z_j(1 - z_j) \quad (9)$$

$$\frac{\partial \hat{y}}{\partial a_j} = w_{j,o} \quad (10)$$

$$\frac{\partial L}{\partial \hat{y}} = -2(y - \hat{y}) \quad (11)$$

And replacing in (5), the expression in this case would be:

$$\frac{\partial L}{\partial w_{i,j}} = \sum_{n}\sum_{m} -2(y - \hat{y})\,.\,w_{j,o}.\,z_j(1 - z_j).\,x_i \quad (12)$$

When the derivatives have been calculated, an optimization algorithm is used to update the weights in the direction of the steepest descent. The most common optimizers used in the literature are Gradient Descent, Stochastic Gradient Descent, Adam, Adadelta, etc., (Kingma & Ba, 2014; Kumari et al., 2023; Ruder, 2016; Zeiler, 2012). There are some other methods used in large-scale problems that use second-order derivative approximations (Bottou et al., 2018). However, the most common optimizer for non-complex cases is gradient descent given in Equation (5) because it is easy to implement, understand and compute. This involves adjusting the weights by subtracting a multiple of the gradient of the error concerning the weights. This multiple is known as the learning rate and controls the size of the update step in each iteration.

$$w_{i,j}^{m+1} = w_{i,j}^{m} - \alpha \left(\frac{\partial L}{\partial w_{i,j}}\right)^{i} \tag{13}$$

where $w_{i,j}^{m+1}$ is the new weight, $w_{i,j}^{m}$ is the weight in the last iteration m, α is the learning rate and $\left(\frac{\partial L}{\partial w_{i,j}}\right)$ is the derivative of the Loss function w.r.t weight.

This process is repeated until the error between the predicted and actual output values is minimized (Kevin P., 2022). The backpropagation algorithm is highly effective in optimizing the weights of an ANN to produce accurate output values and is widely used in various fields such as finance, healthcare, and image recognition (Saranya & Subhashini, 2023).

3.2 Support Vector Machines (SVMs)

SVMs are a class of supervised machine learning algorithms used for classification and regression tasks. It was initially designed to solve pattern recognition problems and Vapnik proposed the method for estimating regressions, constructing multidimensional splines and solving linear operator equations (Vapnik, 1995). The results of applying the support vectors to these problems can be found in (Vapnik et al., 1996). At their core, SVMs are about finding a hyperplane that best separates data into distinct classes. This hyperplane is defined by a linear equation.

Consider a binary classification problem where two classes are given, often labeled as -1 and +1. The goal of SVM is to find a hyperplane that maximizes the margin, the distance between the hyperplane and the nearest data points (support vectors) from each class (Raghavendra. N & Deka, 2014).

The equation of a hyperplane in a feature space is given by:

$$x \cdot w + b = 0 \tag{14}$$

Where:

- x represents the input features.
- w is the weight vector, perpendicular to the hyperplane.
- b is the bias term, which shifts the hyperplane away from the origin.

The decision boundary is defined by this hyperplane. Data points are classified based on which side of the hyperplane they fall: if $x \cdot w + b > 0$, the point belongs to one class, and if $x \cdot w + b < 0$, it belongs to the other class.

The effectiveness of SVMs depends on the linear separability of data. Linearly separable data means that you can draw a straight line (in 2D), a hyperplane (in higher dimensions), that perfectly separates the two classes. Here is an example:

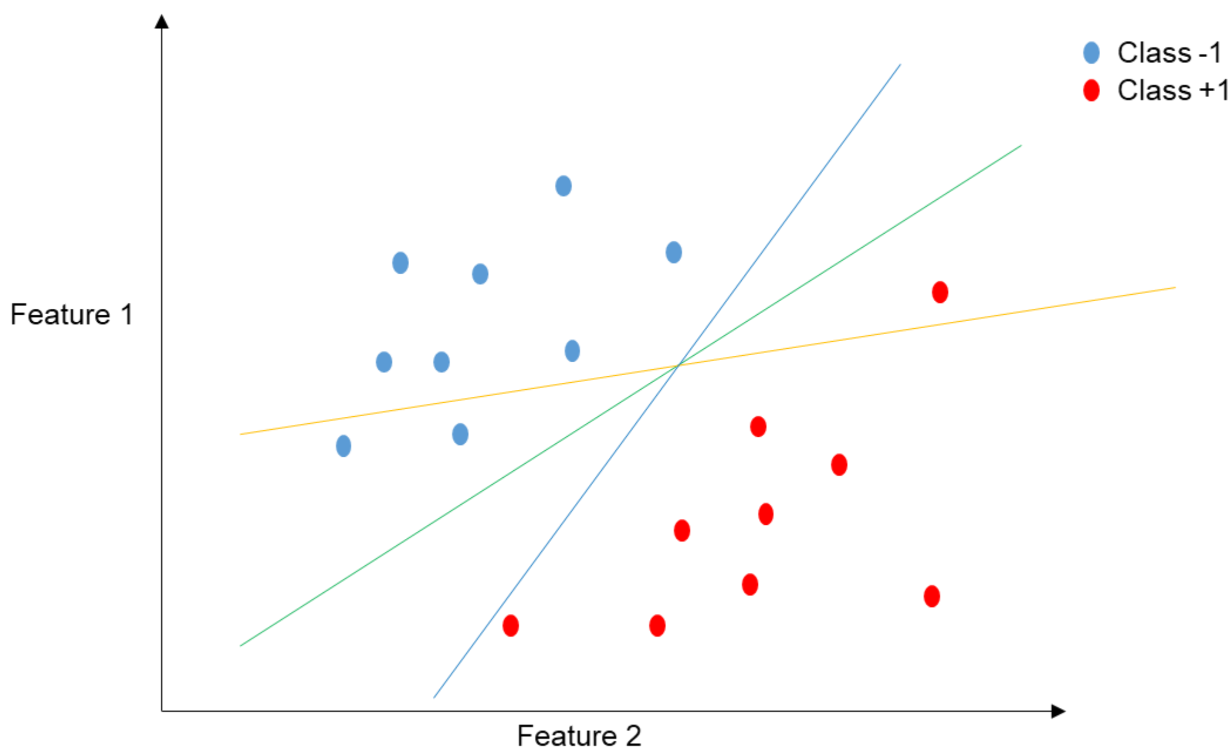

Figure 10 Linear separable example

In Figure 10, three different planes are used to separate both classes, and can be observed that the one that better separate the datasets for class -1 and class +1 is the green one, can be observed that a single straight line can separate the two classes, making this data linearly separable.

However, many real-world datasets are not linearly separable, as showed in Figure 11. In such cases, SVMs can still be applied by transforming the data into a higher-dimensional space using kernel functions. This allows for the creation of a hyperplane in the transformed space that can effectively separate non-linearly separable data.

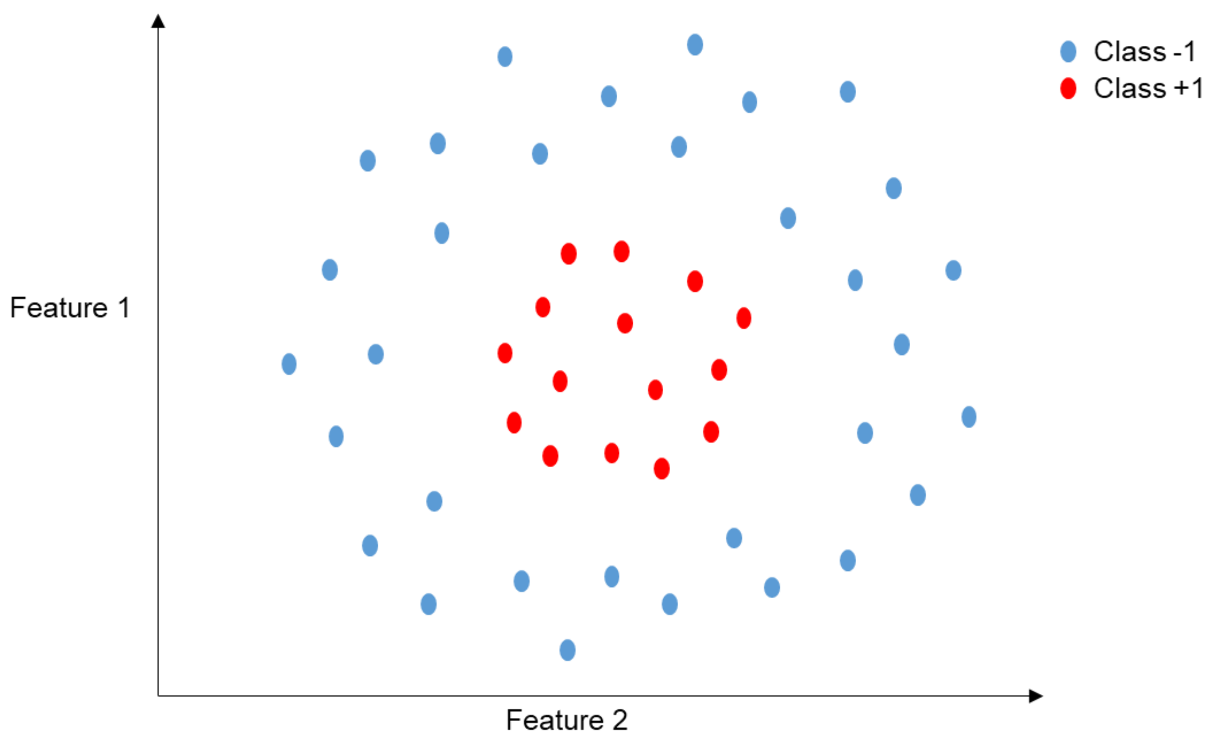

Figure 11 Linear Non-Separable example

The main objective in training an SVM is to find the hyperplane that maximizes the margin while minimizing classification errors. This optimization problem can be formulated as follows:

$$\max_{w} \quad \frac{2}{\|w\|}$$

subject to $\quad y_i(x_i \cdot w + b) \geq 1, \forall i$

Here, y_i is the class label (-1 or +1), x_i is the data point, and $x_i \cdot w + b$ is the decision boundary. The condition $y_i(x_i \cdot w + b) \geq 1$ enforces that all data points are correctly classified with a margin of at least 1.

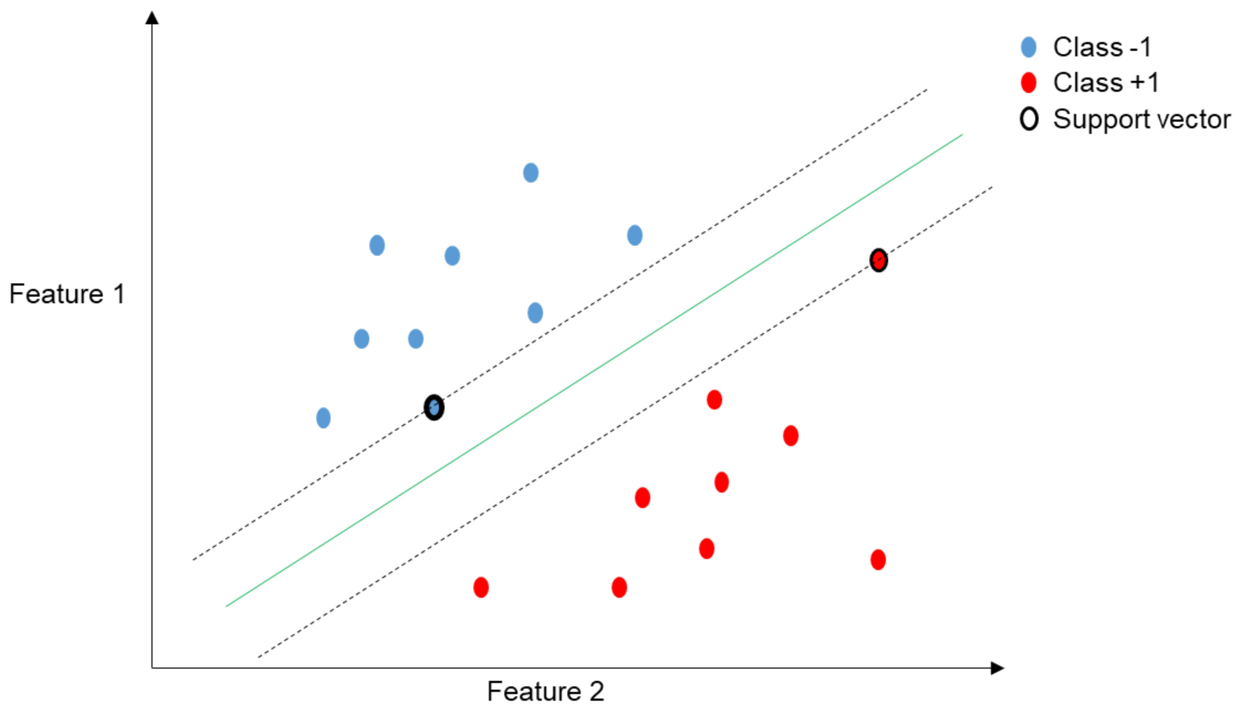

Figure 12 Support vector interpretation

The goal is to find the weight vector w and the bias term b that satisfy this condition while maximizing $\frac{2}{\|w\|}$ which is the same that minimizing $\|w\|^2$. The term $\frac{2}{\|w\|}$ represents the width of the margin, and the optimization seeks to maximize this margin while minimizing classification errors. The data points selected whilst maximizing these margins are known as the support vectors, as is shown in Figure 12.

This kind of optimization problem, where a maximization or minimization of some quadratic function should be done, subject to a number of linear conditions or constraints, is known as quadratic programming (Nocedal & Wright, 2006). This can be solved using techniques like the Lagrange multiplier method (Emami & Omar, 2013; Rockafellar & Wets, 1986), leading to the final SVM model with the optimal hyperplane. The primal formula will be:

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i [y_i(x_i \cdot w + b) - 1] \tag{15}$$

Where α is the Lagrange multiplier.

Here the L_p should be minimized setting the gradient of the Lagrange primal to zero, to find a stationary point for L_p over w and α. Then, the Lagrange partial derivatives are set to zero, $\frac{\partial L_p}{\partial w} = 0, \frac{\partial L_p}{\partial \alpha} = 0.$

Solving, can be found that:

$$\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^{l} \alpha_i y_i x_i = 0 \tag{16}$$

So,

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{17}$$

And,

$$\frac{\partial L_p}{\partial \alpha} = \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{18}$$

Replacing (17) in (15), we have:

$$L_p = \frac{1}{2}\left(\sum_{i=1}^{l} \alpha_i y_i x_i \cdot \sum_{j=1}^{l} \alpha_j y_j x_j\right) - \sum_{i=1}^{l} \alpha_i y_i \left(x_i \cdot \sum_{j=1}^{l} \alpha_j y_j x_j + b\right) + \sum_{i=1}^{l} \alpha_i$$

$$L_p = \frac{1}{2}\left(\sum_{i=1}^{l} \alpha_i y_i x_i \cdot \sum_{j=1}^{l} \alpha_j y_j x_j\right) - \left(\sum_{i=1}^{l} \alpha_i y_i x_i \cdot \sum_{j=1}^{l} \alpha_j y_j x_j\right) - \sum_{i=1}^{l} \alpha_i y_i + \sum_{i=1}^{l} \alpha_i$$

As we found that $\sum_{i=1}^{l} \alpha_i y_i = 0$, we obtain a dual optimization problem:

$$L_p = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{19}$$

Subject to

$$\alpha_i \geq 0$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$

3.2.1 Soft Margin

The concept of “margin” introduced before is known as hard margin, where the goal is to find a hyperplane that perfectly separates the training data into two classes without any classification errors. However, this approach has limitations when dealing with noisy or overlapping data, as real-world data is rarely perfectly separable.

In practical scenarios, it's common for data points to have some degree of noise, outliers, or inherent overlap between classes. Relying on a hard margin approach could result in overfitting the model to the training data and poor generalization to unseen data.

To address these challenges, the concept of a "soft margin" SVM was introduced. In a soft margin SVM, a small amount of classification error is allowed, and the model aims to find a hyperplane that minimizes the sum of the deviations, training errors and maximises the margin width (Cortes & Vapnik, 1995), while considering a user-defined parameter C, which is known as the regularization parameter.

$$\min_{w,C,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$

subject to $\quad y_i(x_i \cdot w + b) \geq 1 - \xi_i$ and $\xi_i \geq 0, \forall i$

where:

- $\frac{1}{2}\|w\|^2$ represents the margin maximization term, like the hard-margin SVM.
- $C\sum_{i=1}^{l}\xi_i$ is the classification error, where ξ_i measures how much each data point is allowed to violate the margin. C is a trade-off parameter that balances the importance of margin maximization and classification errors.

The optimization problem is subject to constraints ensuring that the classification error, represented by ξ_i, is minimized while respecting the user-defined tolerance C. The trade-off in a soft-margin SVM lies in the choice of this parameter.

- A smaller C places more emphasis on maximizing the margin. The model will tolerate more classification errors, allowing for a wider margin. This is suitable for scenarios where the data is relatively clean, and overfitting is a concern.
- A larger C gives more weight to minimizing classification errors, resulting in a narrower margin. The model is less tolerant of errors and will aim to classify training data points correctly. This is useful when dealing with noisy or overlapping data.

The optimal C value is often determined through cross-validation, as it depends on the specific dataset and problem at hand.

3.2.2 Kernel Trick

Linear SVMs are powerful tools for classification tasks when the data is linearly separable, meaning it can be divided into two classes using a straight line (in two dimensions), a hyperplane (in higher dimensions), or a linear decision boundary. However, they have limitations when dealing with non-linearly separable data. These limitations include:

- Inability to Capture Complex Patterns: Linear SVMs struggle to capture complex and non-linear relationships between features and class labels in the data.
- Reduced Classification Accuracy: When data is not linearly separable, using a linear SVM may lead to poor classification accuracy as it tries to fit a linear boundary to inherently non-linear patterns.

The kernel trick is a technique used to overcome the limitations of linear SVMs when dealing with non-linearly separable data. It involves mapping the original feature space into a higher-dimensional space where the data becomes linearly separable. This transformation is achieved by applying a kernel function to the data, which implicitly computes the dot product of data points in the higher-dimensional space without explicitly calculating the transformation.

Generally, it is not commonly known which function transforms the original space into the expanded feature space. Mercer's theorem provides assurance for the existence of such a function, as it dictates certain conditions that must be met (Zhang et al., 2013). Mercer's theorem says: “To be a valid Support Vector Machine kernel, for any finite function $g(x)$ the following integration should always be non-negative for the given kernel function $K(x, y)$” (Vapnik et al., 1996):

$$\iint K(x, y)g(x)g(y) \geq 0$$

Here, $K(x, y)$ computes the inner product in the featured space and is known as the Mercer kernel. Utilizing a positive definite kernel guarantees that the optimization problem will be convex nature and there will be an unique solution (Rakse & Shukla, 2010).

There are several types of kernel functions that can be used in SVMs, each with its own characteristics and properties. These includes linear, polynomial, radial basis functions (RBF) and sigmoid (Raghavendra. N & Deka, 2014; Zhang et al., 2013) :

- ***Linear Kernel:*** The linear kernel (or dot product) is the simplest and is used when the data is already linearly separable. It works well for linear problems and is computationally efficient.

- ***Polynomial Kernel:*** The polynomial kernel raises the dot product of two data points to a certain power, introducing non-linearity. It is effective in capturing moderately complex patterns but can be sensitive to the choice of the polynomial degree.
- ***Radial Basis Function (RBF) Kernel*:** The RBF kernel, also known as the Gaussian kernel, is widely used for its ability to model complex non-linear patterns. It defines similarity between data points based on their Euclidean distance in the transformed space. The RBF kernel has a parameter (gamma) that controls the kernel's flexibility.
- ***Sigmoid Kernel:*** The sigmoid kernel is another non-linear kernel that can capture non-linear patterns. It is based on the hyperbolic tangent function.

These common kernels are shown are shown in Table 5 with their formulas.

Table 5 Common kernel functions

Kernel function	Formula	Annotations
Linear	$K(x_i, x_j) = x_i \cdot x_j$	$x_i \cdot x_j$ is the dot product of the input vectors
Polynomial	$K(x_i, x_j) = (x_i \cdot x_j + c)^d$	c is the constant term and d is the degree of the polynomial
Radial Basis Function (RBF)	$K(x_i, x_j) = \exp(-\gamma \cdot \|x_i - x_j\|^2)$	γ is a parameter that controls the width of the Gaussian function and $\|x_i - x_j\|^2$ is the squared Euclidean distance between the input vectors
Sigmoid	$K(x_i, x_j) = \tanh(\gamma \cdot (x_i \cdot x_j + c))$	c is the constant term

In the literature, some new kernels functions have been proposed, such as KMOD, UKF (Zhang & Wang, 2011), Cauchy (Rakse & Shukla, 2010), Wavelet (Zhang et al., 2013) and ASVMK (Pourbahrami et al., 2023). Also, have been presented kernels based on particle swarm optimized vectors and Bayesian optimizers (Kouziokas, 2020), weighted kernel functions (Wang et al., 2016) and a wide range of families of orthogonal polynomial kernels as Laguerre, Hermite, Jacobi Legendre, Chebyshev and Gegenbauer (Padierna et al., 2018). Limited emphasis will be placed on this aspect as it lies outside the scope of this thesis.

3.2.3 Support Vector Regression

SVMs are employed for both binary and multi-class classification tasks. However, when the aim shifts from predicting discrete class labels to forecasting a continuous target variable, this constitutes a regression problem. In such cases, a specialized version of SVM comes into play. The concept of SVR was originally introduced by Drucker and Vapnik in their paper titled "Support Vector Regression Machines" as the ε-SVR model (Drucker et al., 1996). Here, a different type of loss function is proposed, called the ε-insensitive loss.

$$L_\varepsilon(y,\hat{y}) = \begin{cases} 0 & if\ |y-\hat{y}| < \varepsilon \\ |y-\hat{y}| - \varepsilon & otherwise \end{cases}$$

Where y is the actual value, $\hat{y}$ is the predicted value and ε is the width of the tube.

This function defines a ε tube where if the predicted value is inside the tube, the computed loss is zero, but if the value is outside the tube, the loss is equal to the magnitude of the difference between the predicted value and the ε value given to the tube's radius.

The function is usually written in the following form:

$$J = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} L_\varepsilon(y_i,\hat{y}_i) \tag{20}$$

Some slack variables are introduced to represent the degree to which value lies outside the ε tube.

$$\hat{y}_i \leq y_i + \varepsilon - \xi_i^+ \tag{21}$$

$$\hat{y}_i \geq y_i + \varepsilon - \xi_i^- \tag{22}$$

Given this, the function can be rewritten as follows:

$$J = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i^+ + \xi_i^-) \tag{23}$$

Here, ξ_i^+ and ξ_i^- are zero if the value is located inside the ε-tube. If the predicted value is above the tube, ξ_i^+ is the positive difference between the predicted value and ε, and similarly if the predicted value is below the tube ξ_i^- will not be zero. An interpretation of these parameters is shown in Figure 13.

The constrains are (21), (22), and epsilon positive, which means:

$$\xi_i^+, \xi_i^- \geq 0, \forall i$$

The corresponding Lagrangian equation for the SVM in regression will be:

$$L_p(w, \xi_i^+, \xi_i^-) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{l}\xi_i^+ + \sum_{i=1}^{l}\xi_i^-\right) - \sum_{i=1}^{l}(\mu_i^+\xi_i^+ + \mu_i^-\xi_i^-)$$
$$- \sum_{i=1}^{l}\alpha_i^+[(\hat{y} - y) - (\varepsilon + \xi_i^+)] - \sum_{i=1}^{l}\alpha_i^-[(y - \hat{y}) - (\varepsilon + \xi_i^-)] \tag{24}$$

Where $\alpha_i^-, \alpha_i^+, \mu_i^-$ and μ_i^+ are the Lagrange multipliers

For more details in the final calculation refer to (Drucker et al., 1996). The final form for the prediction is:

$$\hat{y} = \sum_{i=1}^{l}(\alpha_i^- - \alpha_i^+)K(x_i, x) + b \tag{25}$$

Therefore, the goal of ε-SVR is to estimate a function with a constraint that the estimation of each input data point has at most ε deviation from its actual response value, by forming an ε-insensitive tube symmetrically around the estimated function. We can see that the essence of ε-SVR is to perform a linear regression with an ε-insensitive loss function, penalizing predictions that are farther than ε from the desired output. The value of ε is a key factor affecting the flatness of the tube, where a small value leads to a narrow tube; thus, a low tolerance for prediction errors, and a large value leads to a broad tube, thus a high error tolerance.

SVR extends SVMs to regression problems, making them suitable for tasks such as stock price prediction, time series forecasting, and cost estimation (Hamidreza & Morteza, 2019; Henrique et al., 2018; Xu et al., 2019)

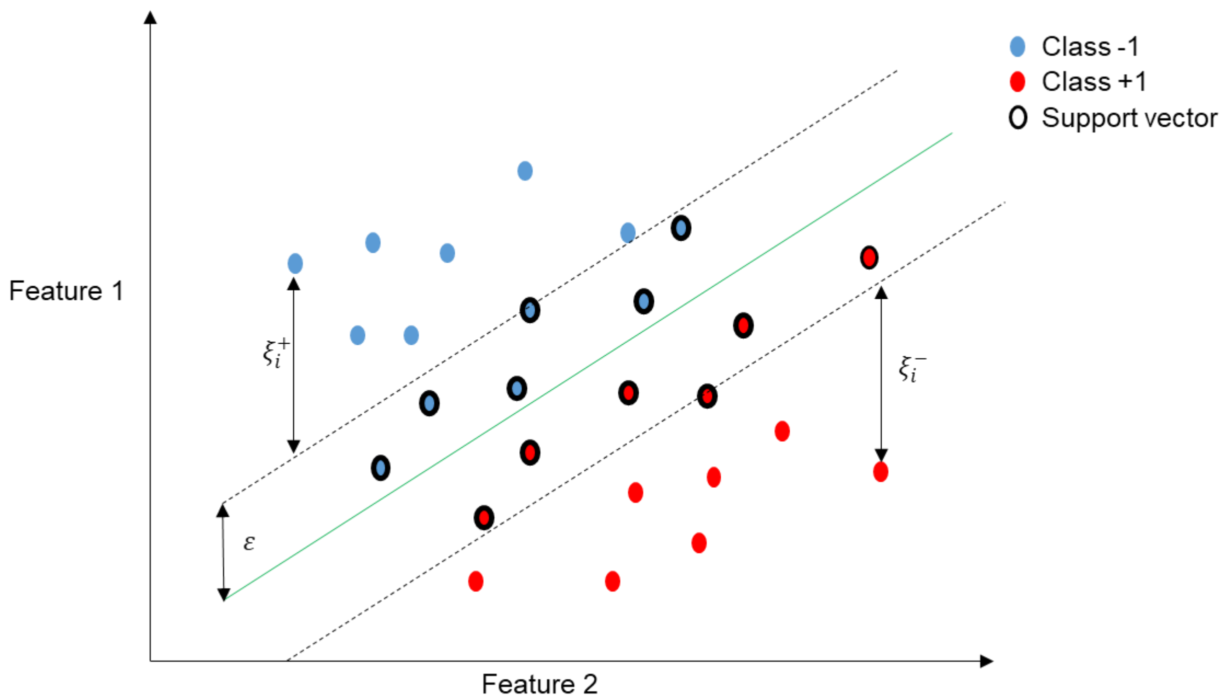

Figure 13 Parameters of SVR

3.3 Hyperparameter selection

In the field of machine learning, the performance and effectiveness of models rely not only on the architecture and training algorithm but also on the careful selection of hyperparameters that are not learned from the data but rather set by the user before training the model. They determine the behavior and characteristics of the model, making their selection a critical aspect of the modelling process. In this regard, the choice of hyperparameters significantly influences the network's ability to learn complex patterns and generalize from the data.

ANN comes with many hyperparameters that need to be properly configured. The hyperparameters of an ANN encompass various aspects, including the network architecture, learning rate, optimizer, regularization techniques, batch size, and number of training epochs, among others (Smith, 2018). Each of these hyperparameters plays a distinct role in shaping the behavior and performance of the network.

The sheer number of hyperparameters and their interdependencies challenge selecting appropriate values. For instance, the choice of optimizer, such as Stochastic Gradient Descent Adam, or RMSprop, may introduce additional hyperparameters specific to that optimizer, such as learning rate schedules or momentum values (Sivaprasad et al., 2020). Similarly, regularization techniques like L1 or L2 regularization introduce hyperparameters like regularization strength (Yu & Zhu, 2020). The complexity of hyperparameter selection increases as we consider different combinations and their potential impact on model performance.

Support Vector Regression (SVR) also entails the configuration of several critical hyperparameters to achieve optimal performance. Much like ANNs, SVR requires careful tuning of these hyperparameters to ensure that the regression model behaves as desired and provides accurate predictions.

SVR hyperparameters encompass various aspects of the model. These include the regularization parameter (often denoted as C), the insensitivity parameter ε, the kernel type (e.g., linear, polynomial, radial basis function), and the associated kernel parameters. Each hyperparameter holds a unique role in shaping the SVR model's behavior and predictive power.

The challenge in SVR, as with ANNs, lies in selecting the appropriate values for these hyperparameters. The choice of the regularization parameter C, for instance, determines the trade-off between fitting the training data precisely and preventing overfitting. A smaller C encourages a wider margin and greater tolerance for errors, while a larger C enforces a narrower margin with fewer deviations from the regression line (Wang et al., 2023).

The parameter ε regulates the width of the ε-insensitive zone, which is employed to accommodate the training data. The choice of ε has implications for the selection of support vectors utilized in building the regression function. When ε is set to a larger value, fewer support vectors are chosen, leading to regression estimates that are more "flat" and less intricate (Cherkassky & Ma, 2004). Consequently, both the values of C and ε influence the model's complexity, although they do so in distinct ways.

Moreover, when dealing with non-linear relationships in the data, selecting the kernel type and its associated parameters becomes crucial. For instance, the radial basis function (RBF) kernel requires setting the kernel width (γ), which influences the flexibility of the SVR model.

Improper selection of hyperparameters can lead to suboptimal model performance. Insufficient regularization may result in overfitting, where the model memorizes the training data but fails to generalize well to unseen data. On the other hand, excessive regularization might cause underfitting, where the model fails to capture important patterns in the data. In the case of ANN, the learning rate, if set too high or too low, can hinder the convergence of the model (Smith, 2018).

Thus, the careful selection of hyperparameters is essential for achieving the right balance and maximizing the model's generalization capabilities.

Given the significance and complexity of hyperparameter selection, researchers and practitioners have developed various strategies to tackle this challenge. Methods range from manual tuning, where experts iteratively adjust hyperparameters based on intuition and experimentation, to automated techniques that leverage computational power and algorithms to search for optimal configurations. These automated methods include grid search, random search, Bayesian optimization, and more advanced techniques that utilize machine learning algorithms for hyperparameter tuning (Bergstra et al., 2011; Bergstra & Bengio, 2012; Claesen & De Moor, 2015; Smith, 2018). Grid search involves exhaustively evaluating the performance of the model for all possible combinations of hyperparameters, while random search involves randomly sampling combinations of hyperparameters. Bayesian optimization and automated hyperparameter tuning are more advanced techniques that use probabilistic models and algorithms to efficiently explore the hyperparameter space (Victoria & Maragatham, 2021).

The choice of hyperparameters could be a specific problem due to the available computational resources. Hyperparameter tuning is an iterative process, and multiple trials may be necessary to identify the optimal configuration. Cross-validation can be employed to evaluate the performance of different hyperparameter settings (Tsamardinos et al., 2015).

Overall, selecting the right hyperparameters is a crucial step in building an effective ANN model, and selecting an appropriate method for hyperparameter tuning can significantly impact the model's performance and generalization ability.

3.4 Evaluation metrics

This study examines the effectiveness of ANN and SVR in estimating underground mining costs by utilizing commonly employed metrics in implementing machine learning models for regression (Michelucci & Venturini, 2023). They can be categorized into accuracy-base and error-base metrics (Dessain, 2022), which include the coefficient of determination (R2) for the accuracy-base metric, and mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean average percentage error (MAPE) for error-base metrics—giving more importance to this last one, given that the main study is to reduce the percentage of deviation in the average of the cost predicted against the actual cost, which is precisely the definition of the MAPE.

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{7}$$

Where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}_i$ is the mean of all the y_i, and n is the number of observations.

4 Case study

The success of implementing any type of machine learning technique heavily depends on the quality and quantity of data provided. One of the most significant advantages of ANN and SVR is their ability to learn and make predictions based on patterns identified in large data sets. However, if the data used for training the network is inaccurate, incomplete, or biased, the resulting predictions may be unreliable or even useless. It is, therefore, crucial to ensure that the data used for the implementation of both algorithms is of high quality, represents the real-world problem being addressed, and is well-prepared for machine learning algorithms. This includes data cleaning, normalization, feature engineering, and data augmentation. In addition, the data must be comprehensive enough to capture all relevant variables and factors that can affect the predicted outcome. In summary, the success of implementing any machine learning technique depends largely on the quality, quantity, and relevance of the data provided.

In this thesis, a case study is presented based on a mine in The Americas. The dataset utilized in this study encompasses a comprehensive summary of the mine's operations over a period of one year. The summary data was derived from ongoing mine activities and cost control measures implemented during this period.

While specific details are safeguarded to ensure the confidentiality of proprietary information, the overall approach to data gathering and analysis adheres to industry standards and best practices. This ensures the integrity and relevance of the findings presented in the case study.

The database used included various variables such as Rock Type, Area (m^2), Drillholes, Drillhole length (feet), Volume per blast (m^3), Jumbo (MH), Scoop (MH) and Dumper (MH). The distribution of each of these variables is shown in Figure 14.

- “Rock type” was categorized from 1 to 5, with 1 being a competent rock and 5 representing a very weak rock.
- “Area” refers to the cross-sectional area given in square meters.
- “Drillholes” refers to the number of drillholes in each working face.
- “Drillhole length” was given in feet.
- “Vol per blast” refers to the volume of a blast given in cubic meters.
- The last three variables, Jumbo, Scoop, and Dumper, represent the machine hours of three main equipment used in the working faces for drilling, mucking, and haulage, respectively.

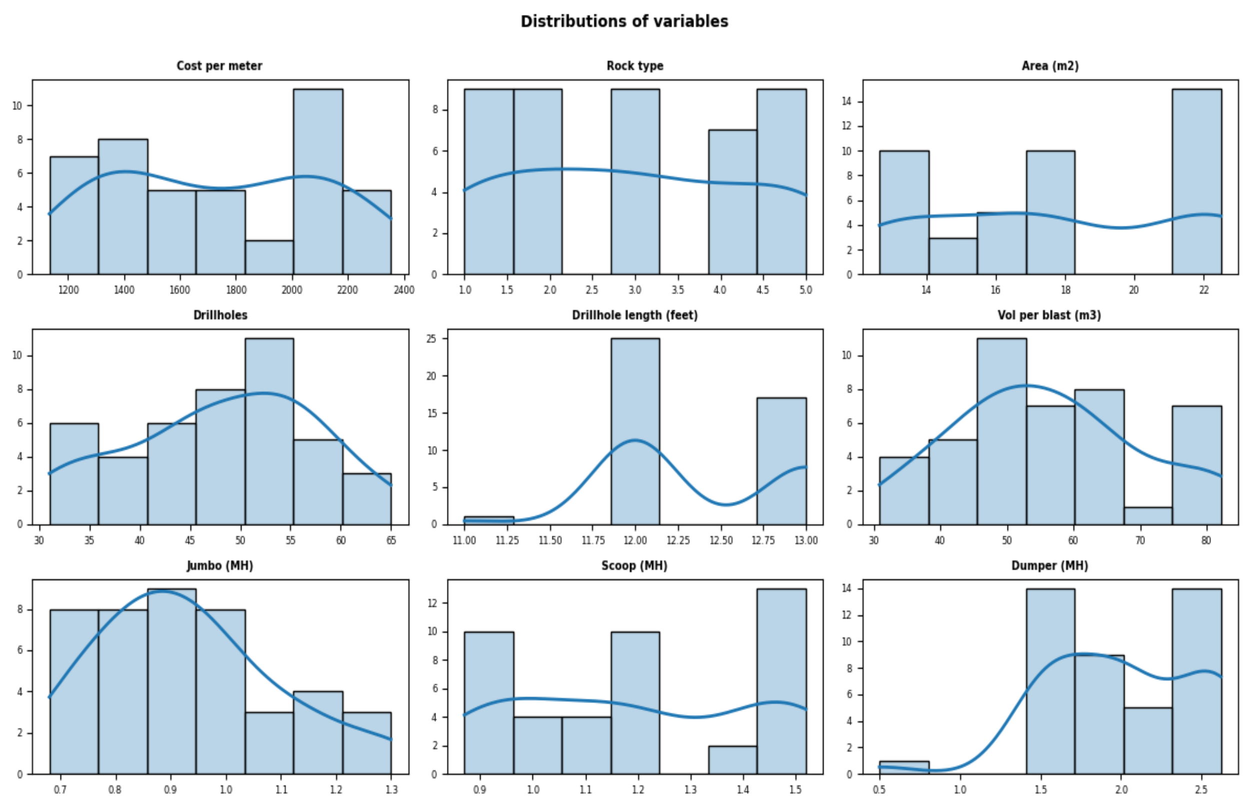

Figure 14 Distributions of variables

As can be seen from the descriptive statistics presented in Table 6, the variables included in the database are of different natures, ranging from operational to rock quality and equipment variables. This wide range of variables allows for a more accurate study, as opposed to only considering variables of the same nature.

Therefore, the data collected from the mine provides valuable insight into the operational performance and rock quality of the mine. Using this data in the machine learning models ensures a more reliable and accurate prediction of the mine's performance.

Table 6 Main descriptive statistics of the features used in the research.

	mean	std	min	max
Cost per meter (USD)	1,735.05	381.39	1,133.59	2,353.89
Rock type	2.95	1.45	1.00	5.00
Area (m^2)	17.77	3.72	12.65	22.50
Drillholes	48.16	9.52	31.00	65.00
Drillhole length (feet)	12.37	0.54	11.00	13.00
Vol per blast (m^3)	56.23	14.10	30.83	82.16
Jumbo (Machine hours)	0.93	0.16	0.68	1.30
Scoop (Machine hours)	1.20	0.23	0.87	1.52
Dumper (Machine hours)	2.01	0.48	0.50	2.62

Measuring the strength of the relationship between random observations is critical in statistics. The correlation coefficient measures the strength of the relationship between variables and typically takes values in the range [-1,1] or [0,1]. The most widely used and well-known correlation coefficient is the Pearson coefficient. It is defined for random variables x and y with finite and positive variances, and it measures the linear relationship between the variables (Edelmann et al., 2021).

A correlation matrix can be used to understand the associations among variables better. This matrix provides a comprehensive summary of the correlation coefficients between all pairs of variables in the dataset. This thesis uses the Pearson correlation coefficient to analyze a correlation matrix presented in Figure 15. The matrix contains information about the correlation between various variables, including operational, rock quality, and equipment variables. By examining the matrix, researchers can gain insights into the relationships between these variables and how they impact the performance of the mine. Overall, using correlation coefficients and correlation matrices is a crucial tool in statistical analysis and can provide valuable insights into the underlying associations among variables.

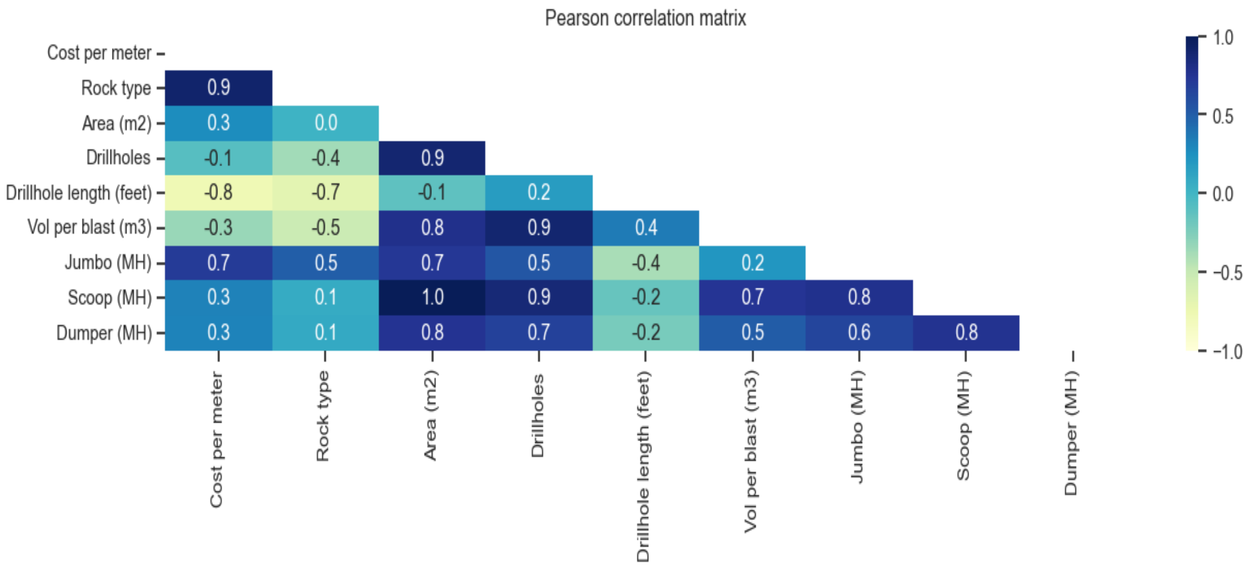

Figure 15 Pearson correlation matrix

To analyze the how is the behavior of the correlation between all the given variables and the cost per meter, the pair plots among all of them are displayed in Figure 16. As can be observed, there are some patterns that can be classified in groups. A fast analysis shows that the variable which is leading the grouping is Rock Type. So, it is decided to set this variable as class for displaying the pair plots.

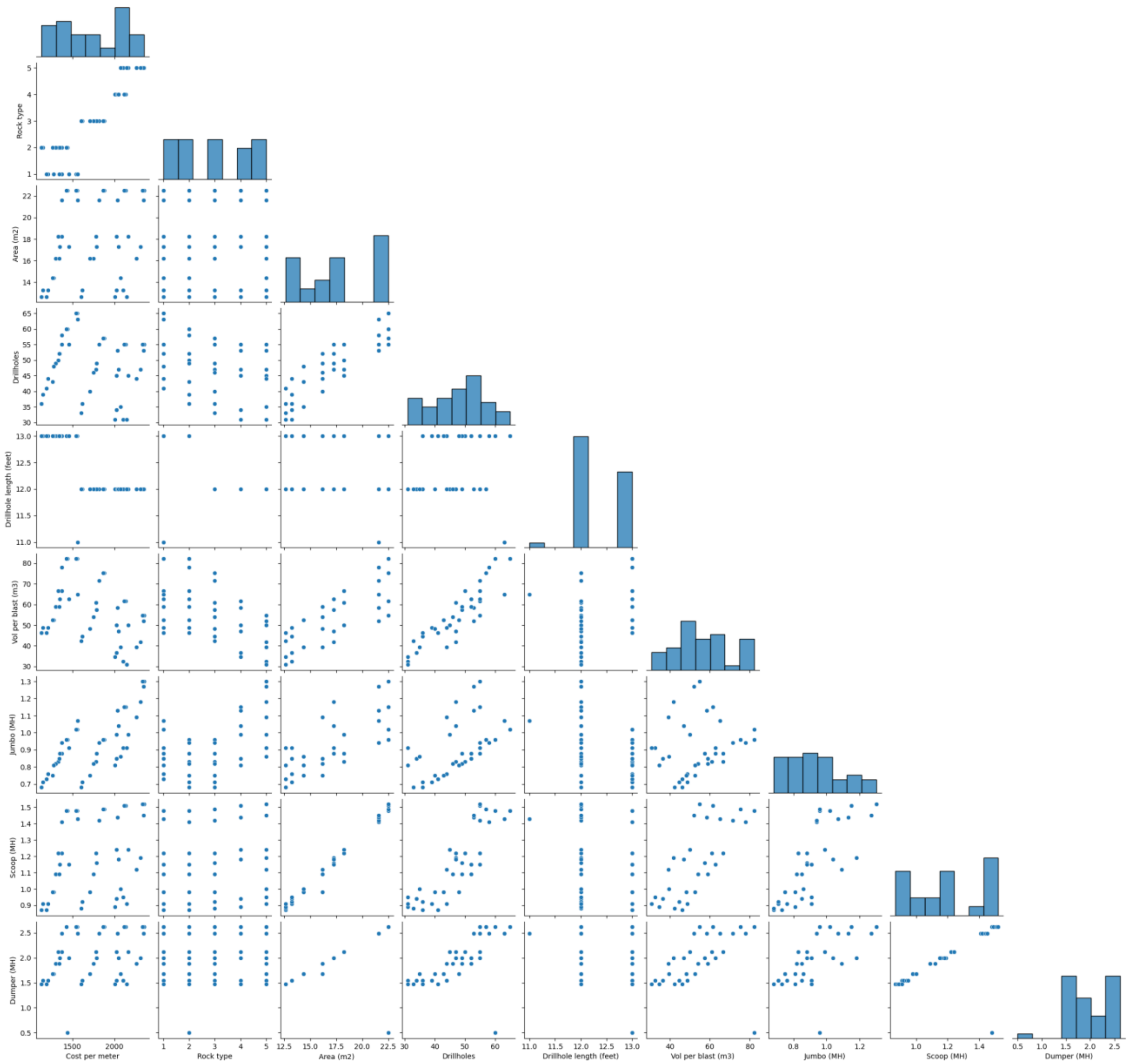

Figure 16 Pair plots among all the variables

Figure 17 shows that the correlations between the cost per meter and most variables are nearly linear, except for the drillhole length variable. However, for all the other variables, it is not straightforward to establish a linear correlation, indicating non-linear characteristics of the input variables. This non-linearity in the input variables makes it challenging for conventional methods to make accurate predictions as they are not equipped to manage such complexities. However, ANN models are well-suited for such scenarios as they can model non-linear relationships and significantly reduce prediction errors.

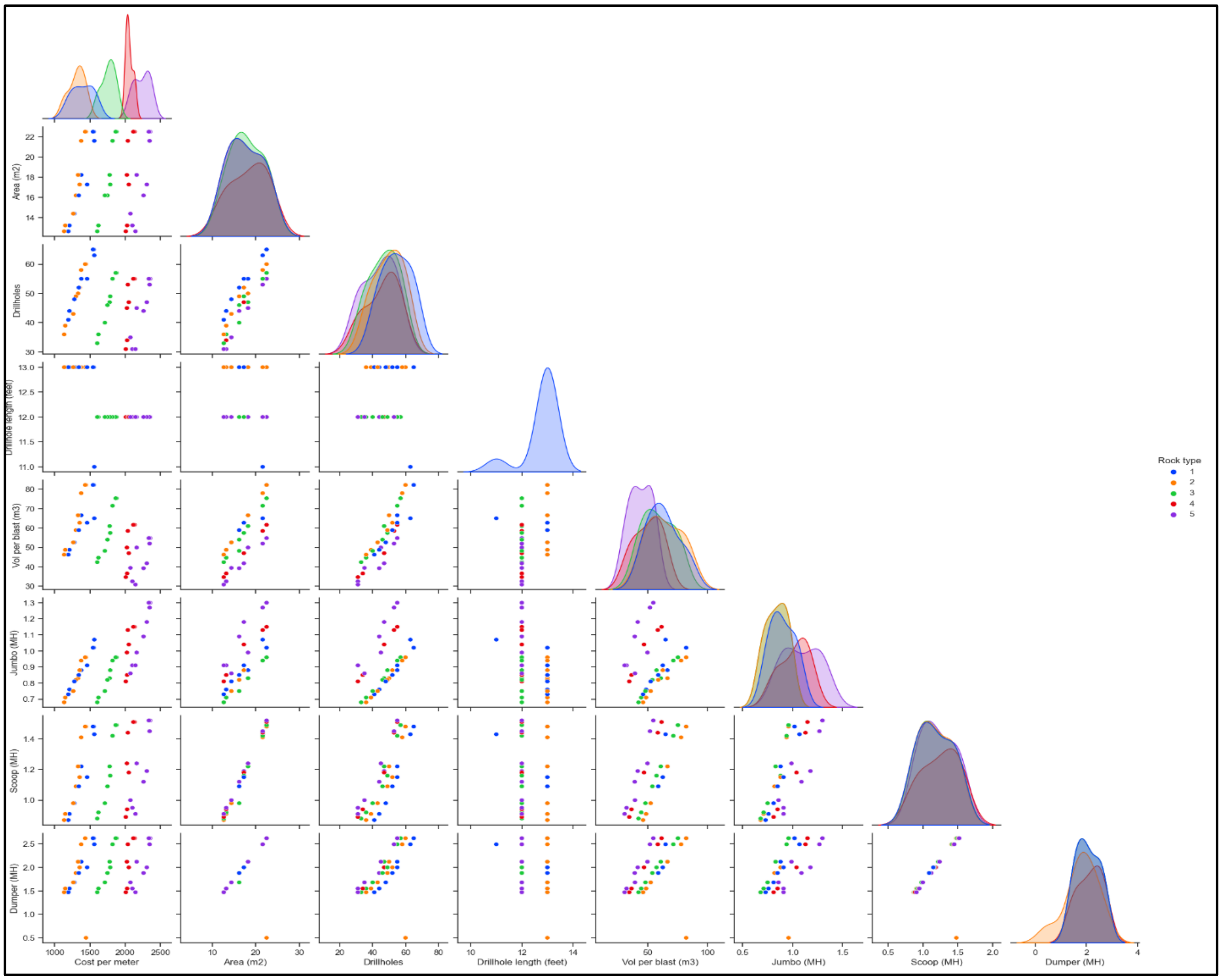

Figure 17 Pair plots among all the variables given having Rock Type as a class.

One of the primary challenges encountered during the implementation process of ANN and SVR is effectively tuning the hyperparameters. As previously discussed, a commonly employed approach called Grid Search entails exhaustively evaluating the model's performance across all possible combinations of hyperparameters. However, it is crucial to define appropriate parameter ranges to avoid excessive computational resources while still exploring optimal parameter values.

This study initially employed a manual tuning approach to establish sensible ranges for the algorithm. This step aimed to prevent the algorithm from searching within an overly broad range, which could consume considerable computational resources without necessarily identifying optimal parameter values. After the manual tuning stage, the parameter ranges presented in Table 7 and Table 8 were selected for the subsequent Grid Search.

Table 7 Parameters considered for ANN hyperparameter tuning with Grid Search

Parameter	Range
Hidden layers	[1, 2]
Hidden neurons	[1,5,10,15,20,25,30,35,40,45,50]
Activation functions	['tanh', 'relu','identity', 'logistic'],
Solver (Optimizer)	['sgd', 'adam']
Type of learning rate	['constant', 'adaptive'],
Learning rate	[0.1, 0.05, 0.01, 0.005, 0.001, 0.0001]
Alpha	[0.0001, 0.001, 0.01, 0.1]

Table 8 Parameters considered for SVR hyperparameter tuning with Grid Search

Parameter	Range
Kernel function	['linear', 'rbf', 'polynomial', 'sigmoid']
C	[1, 1.5, 2, 2.5, 3, 4, 5]
Gamma	[1, 0.1, 0.01, 0.001]
Epsilon	[0.001, 0.01, 0.1, 0.5, 1, 2, 4]

Subsequently, the Grid Search was performed using the predefined parameter ranges. The search systematically evaluated the model’s performance across the provided parameter combinations. The resulting parameter values, determined through the Grid Search, are presented in Table 9 and Table 10.

This systematic approach to hyperparameter tuning enables the identification of optimal parameter values for the ANN and SVR models, enhancing its performance and predictive accuracy. The combination of manual tuning and Grid Search offers a robust and efficient methodology for hyperparameter optimization in the context of the ANN and SVR implementation.

Table 9 ANN hyperparameters selected by Grid Search

Parameter	Value
Hidden layers	1
Hidden neurons	15
Activation functions	relu
Solver (Optimizer)	adam
Type of learning rate	adaptative
Learning rate	0.1
Alpha	0.001

Table 10 SVR hyperparameters selected by Grid Search

Parameter	Value
Kernel function	poly
C	1
Gamma	1
Epsilon	0.01

After determining the optimal hyperparameters, the algorithms are implemented to train the model. A predefined stopping criterion was established to ensure an appropriate stopping point for the training process of ANN. In this study, the criterion was set as a tolerance level of 0.0001 in the R2 Score, requiring no significant improvement observed over ten consecutive iterations. This means that if, after ten continuous iterations, the algorithm does not exhibit an improvement greater than 0.0001 in the R2 Score, the algorithm will consider the current Score and its corresponding loss value as the best and results for the implementation.

During the training iterations for the ANN model, the algorithm calculates the loss value and the validation score. In this study, the obtained loss value was 0.003, indicating the magnitude of the errors between the predicted and actual values. Furthermore, the validation score, with a value of 0.94, assesses the model's performance in capturing the underlying patterns in the data.

The behavior of the loss value and the validation score throughout the iterations is visually depicted in Figure 18 and Figure 19, respectively. These figures provide insights into the convergence and progress of the algorithm during the training process.

By establishing a stopping criterion and closely monitoring the loss value and validation score, the algorithm can effectively determine the optimal point to conclude the training process.

This approach ensures that the model achieves satisfactory accuracy and performance before it is applied to make predictions or perform further analyses.

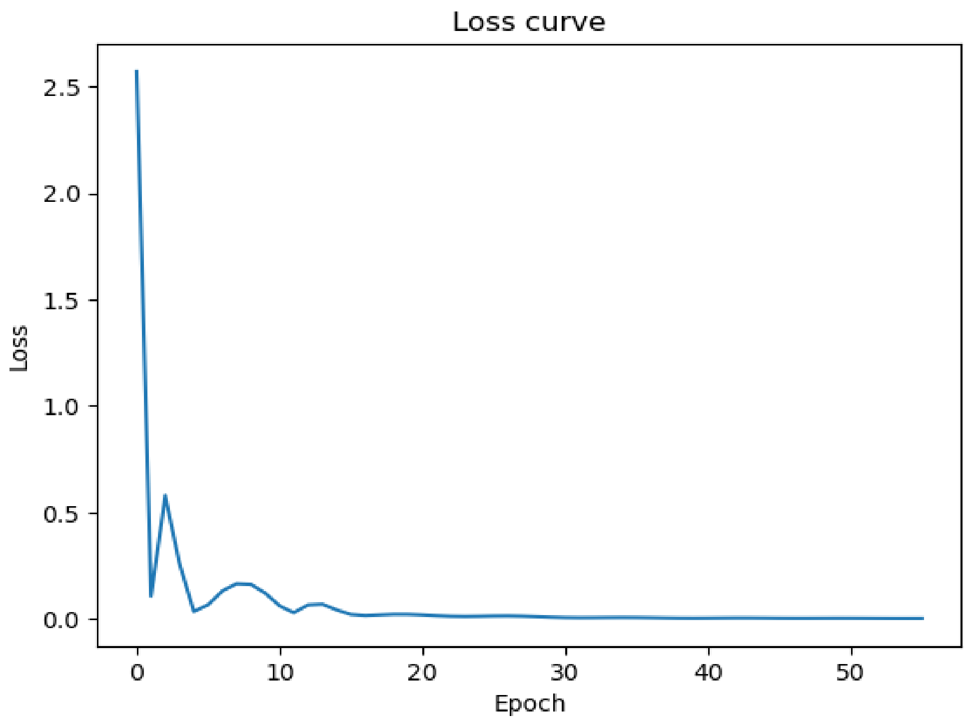

Figure 18 Loss curve in ANN implementation

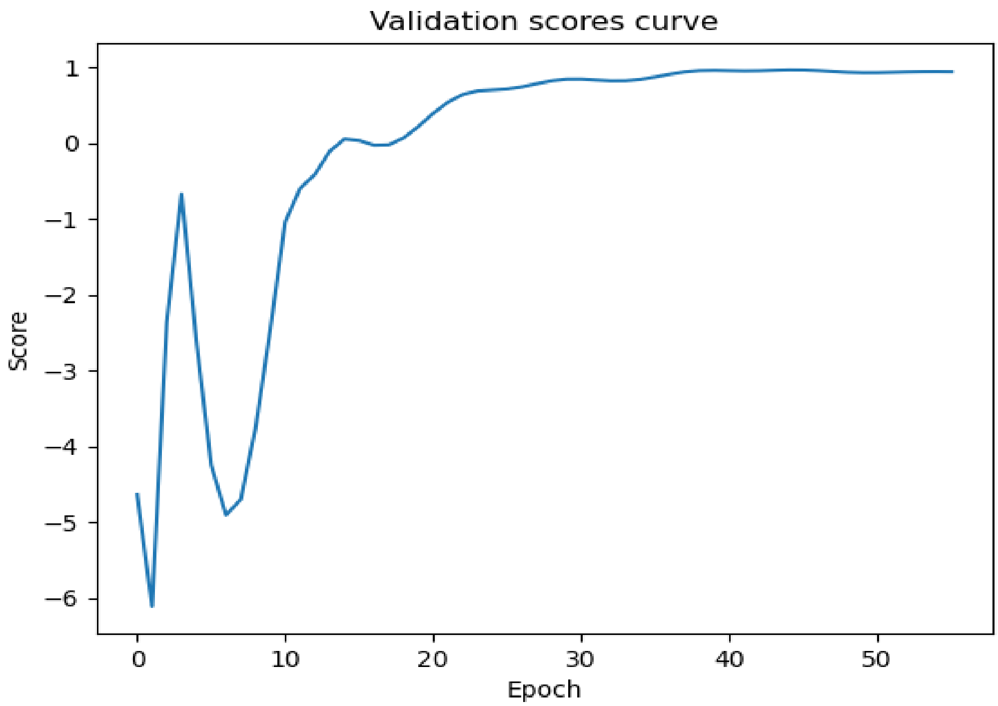

Figure 19 Validation scores curve in ANN implementation

During the evaluation of the models on the test data, a comparison was made between the actual cost and the predicted cost generated by the model. This assessment allows us to gauge the accuracy of the model's predictions. The closeness between the predicted and actual costs is illustrated in Figure 20 and Figure 22, which visually depicts the proximity of the predicted values to the real values.

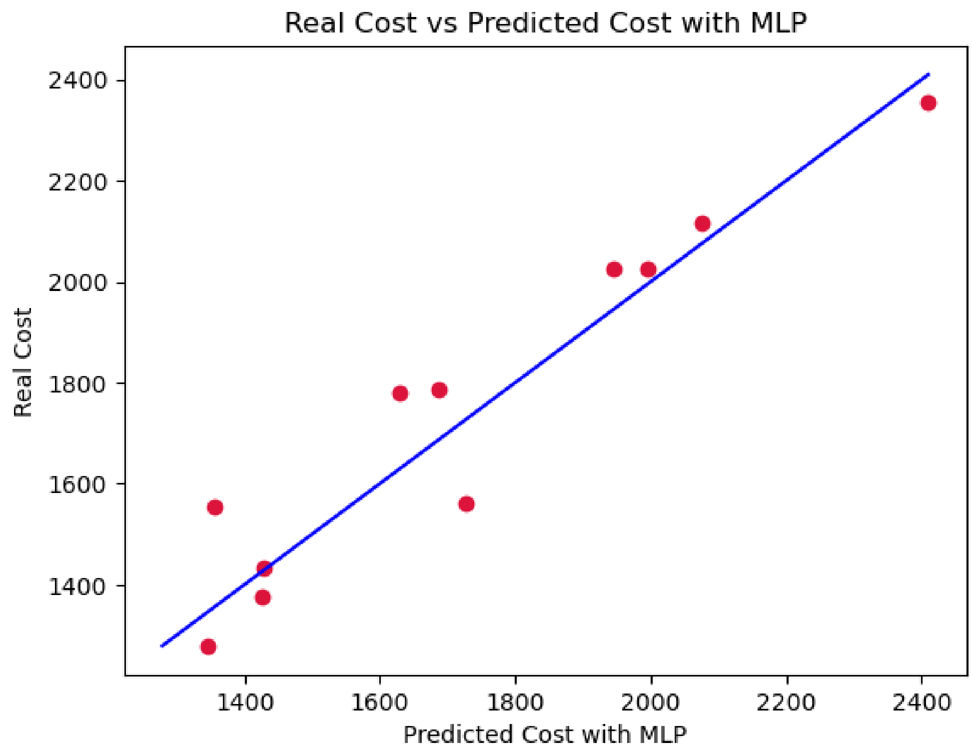

Figure 20 Real cost vs predicted cost with ANN.

To further analyze the performance of the models, Figure 21 and Figure 23 provide a representation of the residual errors between the predicted and actual costs. The maximum deviation observed in the residual errors is 200.15 CAD, translating to an Absolute Percentage Error of 12.87% for the ANN model and 191.95, CAD corresponding to an Absolute Percentage Error of 12.29% for the SVR model. This metric quantifies the relative magnitude of the errors in relation to the actual costs.

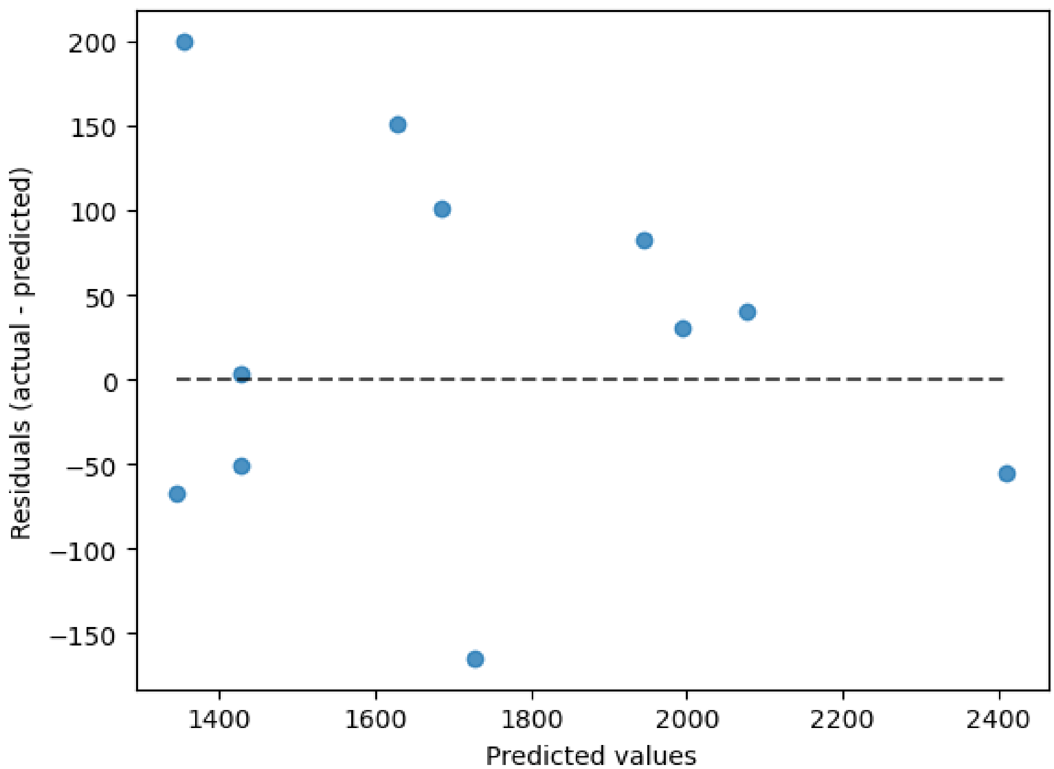

Figure 21 Residual values in the predicted values with ANN

For a comprehensive overview of the residual values and their corresponding Absolute Percentage Error for all predicted values, referring to Table 11 and Table 12. These tables offer a detailed interpretation of the errors and provide insight into the model's performance across multiple cost predictions. The Mean Absolute Percentage Error (MAPE) is calculated as 5.31% and 3.05% for ANN and SVR models, respectively, in this model implementation. This result signifies a better-than-expected level of accuracy, considering that the accepted "accuracy" threshold for cost estimation is typically set at less than 10% (AACE, 2020).

Table 11 Comparison between actual and predicted values ANN.

Actual (USD)	Predicted (USD)	Difference (USD)	Absolute percentage error (%)
2024.19	1994.06	30.13	1.49
1561.66	1726.52	-164.86	10.56
1555.12	1354.97	200.15	12.87
1375.59	1426.69	-51.1	3.71
1432.23	1428.71	3.52	0.25
2116.38	2076.23	40.15	1.9
1278.54	1345.37	-66.83	5.23
2026.97	1943.99	82.98	4.09
2353.89	2409.44	-55.55	2.36
1779.4	1628.52	150.88	8.48
1787.24	1685.82	101.42	5.67

The combination of Figure 20, Figure 21 and Table 11 offers a comprehensive evaluation of the ANN model's performance in predicting costs and the Figure 22, Figure 23 and Table 12 combination shows the same for the SVR model's performance. These metrics and visual representations contribute to a thorough understanding of the model's accuracy and effectiveness in estimating costs within acceptable limits.

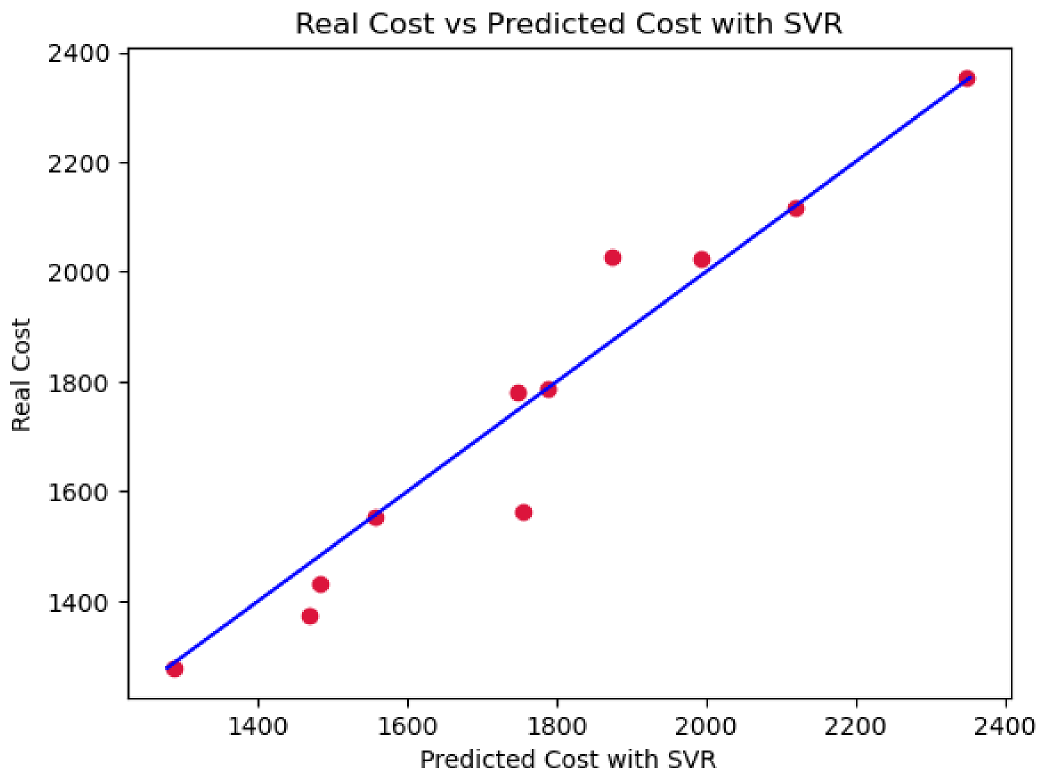

Figure 22 Real cost vs predicted cost with SVR.

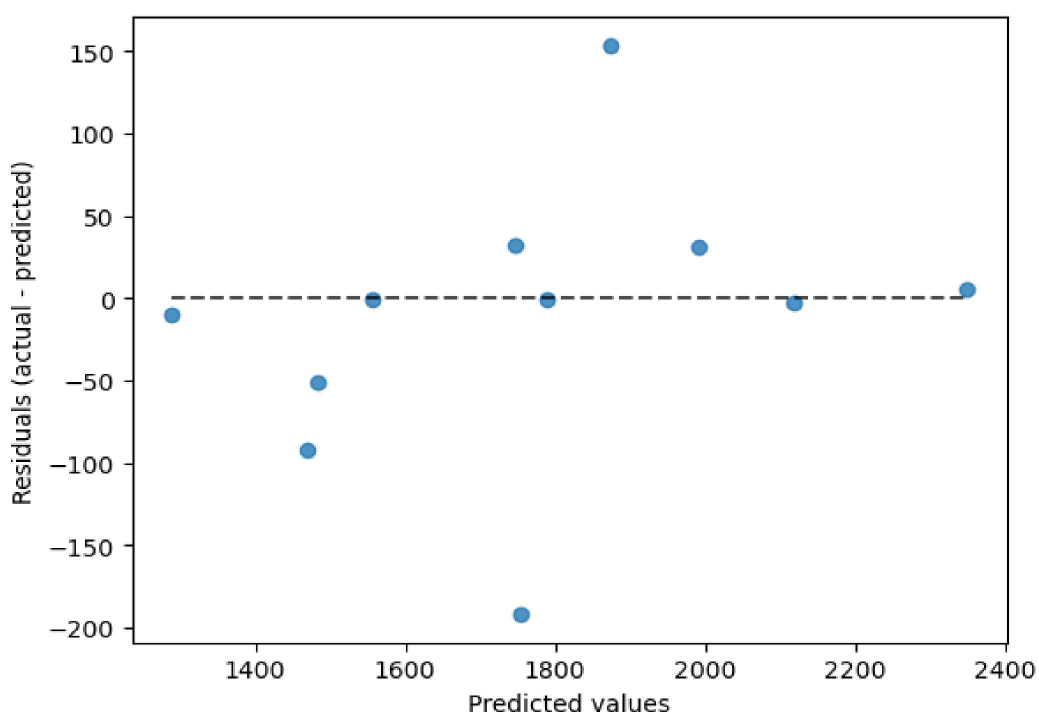

Figure 23 Residual values in the predicted values with SVR

Table 12 Comparison between actual and predicted values SVR.

Actual (USD)	Predicted (USD)	Difference (USD)	Absolute percentage error (%)
1992.51	2024.19	31.68	1.56
1753.61	1561.66	-191.95	12.29
1555.8	1555.12	-0.68	0.04
1468.21	1375.59	-92.62	6.73
1483.47	1432.23	-51.24	3.58
2119.28	2116.38	-2.9	0.14
1288.19	1278.54	-9.65	0.75
1873.31	2026.97	153.66	7.58
2348.93	2353.89	4.96	0.21
1746.84	1779.4	32.56	1.83
1788.15	1787.24	-0.91	0.05

Finally, the Table 13 shows the results with the metrics proposed for evaluation in this thesis. The mean average percentage error is the one we are giving more importance due to its literate definition which is what is used for cost estimation deviations or accuracy expectations and the focus of this study is to reduce the percentage of deviation in the average of the cost predicted against the actual cost. As can be observed, the MAPE obtained in both models fits into the expected values, which are less than 10%.

Table 13 Evaluation metrics results for ANN and SVR

Model	R^2	MSE	RMSE	MAE	MAPE
ANN	0.90	10897.09	104.39	86.14	5.31
SVR	0.94	6713.56	81.94	52.07	3.05

5 Conclusions

Cost estimation plays a crucial role for mining contractors in the bidding process for mining projects, as accurate estimations are essential to avoid detrimental under or overestimations. Based on the findings of this study, it can be concluded that both ANN and SVR serve as reliable techniques for cost estimation, utilizing past information. The achieved MAPE of 5.31% for ANN and 3.05% for SVR, falls within the expected accuracy range for the final stages of cost estimation, while the R^2 score of 0.90 and 0.94, respectively, further validates the effectiveness of both machine learning algorithms.

The selection of hyperparameters in ANN and SVR is of utmost importance, as they significantly influence the network's performance, effectiveness and the degree of error allowed. Properly chosen hyperparameters ensure optimal learning, regularization, and generalization. However, hyperparameter selection poses challenges due to the vast search space and interdependencies among parameters. While grid search was employed in this study, more complex problems may require approaches like Bayesian optimization, considering the availability of computational resources.

Each mining problem should be treated individually, acknowledging the uniqueness of equipment, geometry, geology, and rock differences of mines. This thesis attempts to capture emerging trends in cost estimation, emphasizing the utilization of algorithms like ANN and SVR that not only reproduce the input-output relationship but also consider the interplay among input variables for enhanced prediction.

The outcomes of machine learning techniques should be benchmarked with conventional methods. Experience and expert judgement are still valuable tools, and they can be combined with the proposed techniques in this research. Thus, mining contractors can estimate their costs more accurately.

Future research could greatly benefit from access to more high-quality data that has undergone meticulous cost control, as it serves as valuable input for accurate estimations.

The minerals industry is behind some sectors, such as the construction industry in cost estimation. The cost estimation for underground mining can be more complicated. However, the construction industry has well-developed standards and practices. The minerals industry should investigate their practices to improve their business.

It is proposed to compare ANN and SVR with other machine learning algorithms such as Decision Trees, Random Forest, and maybe different Ensemble methods. This comparison aims to evaluate model performance and assess the proximity of predictions. Furthermore, by exploring alternative algorithms, researchers can gain insights into their strengths and limitations, further advancing the field of mining cost estimation. Also, instead of the Pearson correlation coefficient, the approaches capturing non-linear relationships between variables can be considered (e.g., copulas). Thus, the performance of machine learning techniques can be furthered.

6 References

AACE, I. (2020). AACE International Recommended Practice No. 47R-11: Cost Estimate Classification System - As Applied in Engineering, Procurement, and Construction for the Mining Processing Industries. In: AACE International.

Abu-Doush, I., Ahmed, B., Awadallah, M. A., Al-Betar, M. A., & Rababaah, A. R. (2023). Enhancing Multilayer Perceptron Neural Network using Archive-based Harris Hawks Optimizer to Predict Gold Prices. *Journal of King Saud University - Computer and Information Sciences*, 101557. https://doi.org/https://doi.org/10.1016/j.jksuci.2023.101557

Apicella, A., Donnarumma, F., Isgrò, F., & Prevete, R. (2021). A survey on modern trainable activation functions. *Neural Networks*, *138*, 14-32. https://doi.org/https://doi.org/10.1016/j.neunet.2021.01.026

Awuah-Offei, K., Temeng, V. A., & Al-Hassan, S. (2003). Predicting equipment requirements using SIMAN simulation - a case study. *Mining Technology*, *112*(3), 180-184. https://doi.org/10.1179/037178403225003609

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, *24*.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Bottou, L., Curtis, F. E., & Nocedal, J. (2018). Optimization Methods for Large-Scale Machine Learning. *SIAM Review*, *60*(2), 223-311. https://doi.org/10.1137/16m1080173

Burnett, J. E., & Wampler, B. M. (1998). Unit price contracts: A practical framework for determining competitive bid prices. *Journal of Applied Business Research*, *14*(3), 63-72.

Caputo, A. C., & Pelagagge, P. M. (2008). Parametric and neural methods for cost estimation of process vessels. *International Journal of Production Economics*, *112*(2), 934-954. https://doi.org/https://doi.org/10.1016/j.ijpe.2007.08.002

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, *17*(1), 113-126. https://doi.org/https://doi.org/10.1016/S0893-6080(03)00169-2

Christopher, B. (2006). *Pattern recognition and Machine Learning*. Microsoft Research Ltd. https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/

Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018

Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised Learning. In M. Cord & P. Cunningham (Eds.), *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval* (pp. 21-49). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75171-7_2

Dessain, J. (2022). Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric. *Expert Systems with Applications*, *199*, 116970. https://doi.org/https://doi.org/10.1016/j.eswa.2022.116970

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, *9*.

Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, *503*, 92-108. https://doi.org/https://doi.org/10.1016/j.neucom.2022.06.111

Edelmann, D., Móri, T. F., & Székely, G. J. (2021). On relationships between the Pearson and the distance correlation coefficients. *Statistics & Probability Letters*, *169*, 108960. https://doi.org/https://doi.org/10.1016/j.spl.2020.108960

Emami, M. S., & Omar, K. (2013). A low-cost method for reliable ownership identification of medical images using SVM and Lagrange duality. *Expert Systems with Applications*, *40*(18), 7579-7587. https://doi.org/https://doi.org/10.1016/j.eswa.2013.07.062

Hamidreza, N., & Morteza, O. (2019). Mining capital cost estimation using Support Vector Regression (SVR). *Resources Policy*, *62*, 527-540. https://doi.org/https://doi.org/10.1016/j.resourpol.2018.10.008

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, *4*(3), 183-201. https://doi.org/https://doi.org/10.1016/j.jfds.2018.04.003

Hongquan, G., Hoang, N., Diep-Anh, V., & Xuan-Nam, B. (2019). Forecasting mining capital cost for open-pit mining projects based on artificial neural network approach. *Resources Policy*, *74*, 10. https://doi.org/https://doi.org/10.1016/j.resourpol.2019.101474

Ibbs, C. W., Kwak, Y. H., Ng, T., & Odabasi, A. M. (2003). Project Delivery Systems and Project Change: Quantitative Analysis. *Journal of Construction Engineering and Management*, *129*(4), 382-387. https://doi.org/doi:10.1061/(ASCE)0733-9364(2003)129:4(382)

Kevin P., M. (2022). *Probabilistic Machine Learning: An Introduction*. The MIT Press. https://lccn.loc.gov/2021027430

Khoso, A. R., & Yusof, A. M. (2020). Extended review of contractor selection in construction projects. *Canadian Journal of Civil Engineering*, *47*(7), 771-789. https://doi.org/10.1139/cjce-2019-0258

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kouziokas, G. N. (2020). SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting. *Applied Soft Computing*, *93*, 106410. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106410

Kumari, N., Sharma, N., & Singh, P. (2023). Performance of Optimizers in Text Summarization for News Articles. *Procedia Computer Science*, *218*, 2430-2437. https://doi.org/https://doi.org/10.1016/j.procs.2023.01.218

Layer, A., Brinke, E. T., Houten, F. V., Kals, H., & Haasis, S. (2002). Recent and future trends in cost estimation. *International Journal of Computer Integrated Manufacturing*, *15*(6), 499-510. https://doi.org/10.1080/09511920210143372

Liu, B. (2011). Supervised Learning. In B. Liu (Ed.), *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data* (pp. 63-132). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19460-3_3

Michelucci, U., & Venturini, F. (2023). New metric formulas that include measurement errors in machine learning for natural sciences. *Expert Systems with Applications*, *224*, 120013. https://doi.org/https://doi.org/10.1016/j.eswa.2023.120013

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b*, *4*, 51-62.

Nocedal, J., & Wright, S. J. (2006). Quadratic programming. *Numerical optimization*, 448-492.

Olanrewaju, A., Bong, M. Z. X., & Preece, C. (2022). Establishment of pre-qualification criteria for the selection of subcontractors by the prime constructors for building projects. *Journal of Building Engineering*, *45*, 103644. https://doi.org/https://doi.org/10.1016/j.jobe.2021.103644

Padierna, L. C., Carpio, M., Rojas-Domínguez, A., Puga, H., & Fraire, H. (2018). A novel formulation of orthogonal polynomial kernel functions for SVM classifiers: The Gegenbauer family. *Pattern Recognition*, *84*, 211-225. https://doi.org/https://doi.org/10.1016/j.patcog.2018.07.010

Picornell, M., Pellicer, E., Torres-Machí, C., & Sutrisna, M. (2017). Implementation of Earned Value Management in Unit-Price Payment Contracts. *Journal of Management in Engineering*, *33*(3), 06016001. https://doi.org/doi:10.1061/(ASCE)ME.1943-5479.0000500

PMI. (2017). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)* (Sixth ed.). Project Management Institute, Inc. https://www.pmi.org/

Pourbahrami, S., Balafar, M. A., & Khanli, L. M. (2023). ASVMK: A novel SVMs Kernel based on Apollonius function and density peak clustering. *Engineering Applications of Artificial Intelligence*, *126*, 106704. https://doi.org/https://doi.org/10.1016/j.engappai.2023.106704

Raghavendra. N, S., & Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, *19*, 372-386. https://doi.org/https://doi.org/10.1016/j.asoc.2014.02.002

Rakse, S. K., & Shukla, S. (2010). Spam classification using new kernel function in support vector machine. *International Journal on Computer Science and Engineering*, *2*(5), 1819.

Ran, W., Vahid, A., Clara Man, C., Shu-Chien, H., & Chia-Jung, L. (2022). Assessing effects of economic factors on construction cost estimation using deep neural networks. *Automation in Construction*, *134*. https://doi.org/https://doi.org/10.1016/j.autcon.2021.104080

Rockafellar, R. T., & Wets, R. J.-B. (1986). A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming. In A. Prékopa & R. J. B. Wets (Eds.), *Stochastic Programming 84 Part II* (pp. 63-93). Springer Berlin Heidelberg. https://doi.org/10.1007/BFb0121126

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210-229. https://doi.org/10.1147/rd.33.0210

Sanaz Tayefeh, H., Omid Mahdi, E., & Harleen, K. (2020). Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. *SN Applied Sciences*, *2*. https://doi.org/https://doi.org/10.1007/s42452-020-03497-1

Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, *7*, 100230. https://doi.org/https://doi.org/10.1016/j.dajour.2023.100230

Singh, B., Patel, S., Vijayvargiya, A., & Kumar, R. (2023). Analyzing the impact of activation functions on the performance of the data-driven gait model. *Results in Engineering*, *18*, 101029. https://doi.org/https://doi.org/10.1016/j.rineng.2023.101029

Sivaprasad, P. T., Mai, F., Vogels, T., Jaggi, M., & Fleuret, F. (2020). *Optimizer Benchmarking Needs to Account for Hyperparameter Tuning* Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research. https://proceedings.mlr.press/v119/sivaprasad20a.html

SME. (2011). *SME Mining Engineering Handbook* (P. Darling, Ed. Third ed.). Society for Mining, Metallurgy, and Exploration, Inc.

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.

Suglo, R. (2010). Contract Mining versus Owner Mining–The Way Forward. *Ghana Mining Journal*, *11*, 61-68.

T. Alan O'Hara, S. C. S. (1992). Chapter 6.3 Costs and Cost Estimation. In *SME Mining Engineering Handbook* (Vol. Vol. 1).

Tao, P., Cheng, J., & Chen, L. (2022). Brain-inspired chaotic backpropagation for MLP. *Neural Networks*, *155*, 1-13. https://doi.org/https://doi.org/10.1016/j.neunet.2022.08.004

Tsamardinos, I., Rakhshani, A., & Lagani, V. (2015). Performance-estimation properties of cross-validation-based protocols with simultaneous hyper-parameter optimization. *International Journal on Artificial Intelligence Tools*, *24*(05), 1540023.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer New York, NY. https://doi.org/https://doi.org/10.1007/978-1-4757-3264-1

Vapnik, V., Golowich, S., & Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Advances in neural information processing systems*, *9*.

Victoria, A. H., & Maragatham, G. (2021). Automatic tuning of hyperparameters using Bayesian optimization. *Evolving Systems*, *12*(1), 217-223. https://doi.org/10.1007/s12530-020-09345-2

Wang, X.-Y., Liang, L.-L., Li, W.-Y., Li, D.-M., & Yang, H.-Y. (2016). A new SVM-based relevance feedback image retrieval using probabilistic feature and weighted kernel function. *Journal of Visual Communication and Image Representation*, *38*, 256-275. https://doi.org/https://doi.org/10.1016/j.jvcir.2016.03.008

Wang, Y.-G., Wu, J., Hu, Z.-H., & McLachlan, G. J. (2023). A new algorithm for support vector regression with automatic selection of hyperparameters. *Pattern Recognition*, *133*, 108989. https://doi.org/https://doi.org/10.1016/j.patcog.2022.108989

Watt, D. J., Kayis, B., & Willey, K. (2010). The relative importance of tender evaluation and contractor selection criteria. *International Journal of Project Management*, *28*(1), 51-60. https://doi.org/https://doi.org/10.1016/j.ijproman.2009.04.003

Wiering, M. A., & Van Otterlo, M. (2012). Reinforcement learning. *Adaptation, learning, and optimization*, *12*(3), 729.

Xu, S., Chan, H. K., & Zhang, T. (2019). Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. *Transportation Research Part E: Logistics and Transportation Review*, *122*, 169-180. https://doi.org/https://doi.org/10.1016/j.tre.2018.12.005

Yoosefzadeh Najafabadi, M., Hesami, M., & Eskandari, M. (2023). Machine Learning-Assisted Approaches in Modernized Plant Breeding Programs. *Genes*, *14*(4), 777. https://www.mdpi.com/2073-4425/14/4/777

Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. *arXiv preprint arXiv:2003.05689*.

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

Zhang, R., & Wang, W. (2011). Facilitating the applications of support vector machine by using a new kernel [Article]. *Expert Systems with Applications*, *38*(11), 14225-14230. https://doi.org/10.1016/j.eswa.2011.04.235

Zhang, X., Liu, X., & Wang, Z. J. (2013). Evaluation of a set of new ORF kernel functions of SVM for speech recognition. *Engineering Applications of Artificial Intelligence*, *26*(10), 2574-2580. https://doi.org/https://doi.org/10.1016/j.engappai.2013.04.008