MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

https://mjs.um.edu.my

Predictive Lead Scoring Models: A Contemporary Review of Identified Gaps and Future Directions

Jia Yee Lee[1a]*, Chi Wee Tan[2a], Chuk Fong Ho[3b], and Noor Aida Husaini[4a]

Abstract: Lead scoring plays a pivotal role in modern marketing strategies, aiding businesses in prioritising and optimising their sales efforts. This review paper provides a contemporary analysis of lead scoring models, focusing on their development, application, and effectiveness across Machine Learning algorithms. The paper begins by introducing the concept and evolution of lead scoring, highlighting the shift in how previous researchers developed lead scoring models, from traditional Machine Learning algorithms to more advanced approaches, such as Ensemble Learning, Neural Networks, and Deep Learning algorithms. It also addresses existing gaps in current models and clarifies the misconceptions about the definition of "lead scoring". In addition, the paper reviews the methodologies used in lead scoring, including data collection, feature selection, and the algorithms applied. Finally, it examines emerging trends and challenges in lead scoring, offering an insightful overview of the current landscape and suggesting future research directions to unlock new opportunities for businesses to enhance their marketing strategies.

Keywords: Lead Scoring, Machine Learning, Artificial Neural Networks, Deep Learning, Social Media

1. Introduction

In today's highly competitive business environment, effectively managing and prioritising potential customers is essential for maximising sales efficiency and refining marketing strategies (Morgado, 2018). Lead scoring, a systematic approach to ranking prospects based on their likelihood of conversion (Rodrigues, 2020), has become crucial for businesses aiming to optimise their sales processes and resource allocation (Ceccatelli, 2023; Lindahl et al., 2017; K. K. Sharma et al., 2023; Sundman, 2021).

Traditionally, manual lead scoring relied heavily on heuristic methods, drawing on the intuition and experience of sales teams (Duncan & Elkan, 2015; Nygård & Mezei, 2020). However, with the advent of big data and advancements in Machine Learning (ML), lead scoring has evolved from a traditional approach to a predictive one. Modern predictive models yield more accurate and actionable insights beyond human capability (Wu et al., 2024) by leveraging large customer data and advanced ML algorithms, such as Neural Networks (NN) (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Espadinha-Cruz et al., 2021; Nygård & Mezei, 2020), Deep Neural Networks (DNN) (Chaudhuri et al., 2021), Feedforward Neural Networks (FNN) (Puravankara & Narendra Babu, 2020), and Recurrent Neural Network (RNN) (Giorcelli, 2019). Thus, this review aims to synthesise recent advancements and provide a road map for future research and practical applications.

Specifically, the review paper aims to achieve four key objectives. First, it seeks to classify and compare ML algorithms and feature engineering techniques used in building lead scoring models. Second, it analyses the publicly available "X Online Education" dataset to assess its applicability in this domain. Third, it evaluates the performance metrics and limitations of current lead scoring models. Finally, it identifies emerging trends and proposes actionable future research directions to advance this field.

The paper is organised as follows: Section 2 provides an overview of lead scoring models, discussing the concepts and data used in building these models. Section 3 introduces fundamental concepts of ML algorithms, which are categorised into traditional and advanced ML types. Section 4 offers an introduction to feature engineering techniques. Section 5 analyses patterns and findings from existing lead scoring-related work, focusing on trends in ML types and algorithms, insights from the "X Online Education" dataset, and current applications of performance metrics and feature engineering methods. Section 6 examines the

Authors information:
[a]Department of Computer Science and Data Science Faculty of Computing and Information Technology (FOCS), Tunku Abdul Rahman University of Management and Technology (TAR UMT), Jalan Genting Kelang, Setapak, 53300 Wilayah Persekutuan Kuala Lumpur, MALAYSIA.
E-mail: jylee-wm19@student.tarc.edu.my[1], chiwee@tarc.edu.my[2], nooraida@tarc.edu.my[4]
[b]Department of Software Engineering and Technology, Faculty of Computing and Information Technology (FOCS), Tunku Abdul Rahman University of Management and Technology (TAR UMT), Jalan Genting Kelang, Setapak, 53300 Wilayah Persekutuan Kuala Lumpur, MALAYSIA.
E-mail: cfho@tarc.edu.my[3]

*Corresponding Author:
jylee-wm19@student.tarc.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

challenges in the field and suggests potential future research directions to guide upcoming studies. Finally, Section 7 presents the conclusions of the review.

2. Lead scoring models

2.1 Lead

A lead, also referred to as a business lead or sales lead, is an individual who has shown interest in an organisation's brand, products, or services (Waters, 2022). Sales or marketing teams often target leads as potential customers, as they are the contact persons that the teams aim to pitch to (Fahad, 2017). Leads can be ascertained through various channels, including emails, phone calls, search engine behaviours, in-person visits, forms, or surveys (Niemi, 2022).

2.2 Traditional Lead Scoring

Once leads are collected, scores that indicate whether they are potential customers or not are calculated and assigned to them (Benhaddou & Leray, 2017). Manual lead scoring, also known as traditional lead scoring, is the process of identifying potential leads based on certain criteria determined by an organisation's sales or marketing department, with the usage of a scoring matrix (Ayaz, 2023). A scoring matrix is a tool used to evaluate and rank potential leads based on criteria, such as demographic information and behavioural data.

Figure 1 provides an example of how a scoring matrix is constructed, which may vary between organisations. Each criterion is assigned a weight, and leads are scored based on how well they meet these criteria, with points allocated for specific attributes. This results in a total score that helps rank the leads in order of priority (Nygård & Mezei, 2020). Typically, lead scores categorise prospects as hot (high scores, immediate follow-up), warm (moderate scores, needs nurturing), or cold (low scores, low priority).

Activity	Points
Form/Landing Page Submission	+ 5
Submitted "Contact Me" Form	+25
Received an Email	0
Email Open	+1
Email Clickthrough	+3
Registered for Webinar	+3
Attended Webinar	+10
Downloaded a Document	+5
Visited a Landing Page	+2
Unsubscribed from Newsletter	-2
Watched a Demo	+8
Contact is a CXO	+5
Visited Trade Show Booth	+3
Visited Pricing Page	+10

Figure 1. Scoring Matrix (Autopilot, 2016).

However, manual lead scoring has its limitations. First, it risks missing predictive signals of purchasing behaviour since data is often captured by salesmen or marketers who may not always record all relevant interactions or insights. For instance, a lead might repeatedly view pricing pages or inquire during sales calls, which are not likely to be recorded in manual systems. Another example occurs when salesmen or marketers prioritise capturing immediate customer responses and overlook subtle behavioural cues, such as browsing patterns or social media interactions, which can be crucial for accurate lead scoring. This incomplete data can result in a skewed understanding of a lead's true potential, potentially leading to less effective targeting and missed opportunities. Second, manual lead scoring is time-intensive and cannot process the large data volumes needed for statistically significant insights (Pereira, 2021). Thus, these limitations have driven the adoption of predictive lead scoring, which is achieved by using ML algorithms to automate the lead scoring process (Etminan, 2021).

2.3 Modern Lead Scoring

Figure 2 illustrates the process of automated lead scoring.

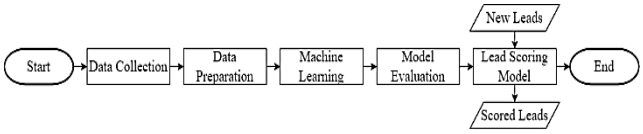

Figure 2. Process of Automated Lead Scoring.

The initial phase involves data collection, where labelled data is gathered specifically for training the predictive model. This is different from the typical data collection done in real-world situations, which often involves unlabelled data intended for lead conversion purposes. The second phase is data preparation, which includes data cleaning, transformation, and feature selection to ensure the data is suitable for modelling. The third phase involves implementing ML algorithms to learn lead conversion patterns from the training data. If available, validation data is used in this phase to fine-tune the model's parameters, ensuring that the model neither overfits nor underfits the training data. In the fourth phase, model evaluation is conducted to assess how well the model generalises to unseen data by using test data that was not included during training or validation. Ultimately, a lead scoring model is built with the capability to predict the likelihood of lead conversion by assigning scores to new leads. Ultimately, a lead scoring model is built with the capability to predict the likelihood of lead conversion by assigning scores to new leads.

2.4 Lead Scoring Datasets

The data used in predictive lead modelling can be categorised into two (2) main types: social media data and non-social media data. Table 1 summarises the characteristics of the fourteen datasets utilised across sixteen studies, with some studies employing the same datasets.

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Table 1. Dataset Characteristics Table.

Characteristics	Data Type	Description	Related Work
Structured, Research records	Non-social media	Public Dataset entitled "X Online Education" obtained from websites, forms and past referrals.	(Jadli et al., 2022; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Yim et al., 2024)
Structured, Sales and Marketing records	Non-social media	Sales and Marketing data stored in Salesforce from 2018 to 2019.	(Ayaz, 2023)
Structured, Marketing records	Non-social media	B2B Dataset was collected as proprietary data from a company located in the western United States B2C Dataset was extracted from a France-based company, Criteo AI Lab's website.	(Bhatta, 2022)
Structured, Web browsing records	Non-social media	Data collected from Google Analytics API.	(Etminan, 2021)
Structured, Web browsing records	Non-social media	Web browsing data collected from an online e-commerce platform based in Germany and Belgium.	(Chaudhuri et al., 2021)
Structured, Marketing records	Non-social media	Data from a telecommunications company across channels (telemarketing, shop, website).	(Espadinha-Cruz et al., 2021)
Structured, Marketing records	Non-social media	Data from HUUB's Customer Relationship Management platforms, HubSpot.	(Pereira, 2021)
Structured, Marketing records	Social media and Non-social media	14 months of lead conversion data from different channels.	(Karthik et al., 2020)
Structured, Web browsing records	Non-social media	Real-life data from an international company includes insights on potential leads in Finland gathered from internal systems, website visits, email sends, opens, click-throughs, and form submissions.	(Nygård & Mezei, 2020)
Structured, Sales records	Non-social media	Customer's sales data over the 3 months from the grocery superstore's billing system named "Taradin Super Shop".	(Choudhury & Nur, 2019)
Structured, Web browsing records	Non-social media	An imbalanced dataset of website visitors' behaviour generated from Google Analytics 360 by Greenchoice for the entire year of 2018.	(Swelsen, 2019)
Structured, Form data	Non-social media	Self-reported information input by the user on the lead form.	(Giorcelli, 2019)
Unstructured, News records	Non-social media	News about different companies on the Internet.	(Benhaddou & Leray, 2017)

Based on Table 1, non-social media data refers to data sources that are not derived from social media platforms. This category includes, but is not limited to, data collected from websites, responses gathered through forms or surveys, and records maintained within databases. Most non-social media datasets are structured and consist of records related to research, sales, marketing, web browsing, forms, and news. Notably, only one non-social media dataset in the context of lead scoring work is unstructured, as presented in the work of Benhaddou & Leray (2017), which contains news data about different companies collected from the Internet. Outside the context of lead scoring, numerous examples of unstructured non-social media data exist, including textual clinical notes (Goh et al., 2021), call transcriptions (Vo et al., 2021), and customer feedback (De Haan & Menichelli, 2019).

Social media data, on the other hand, encompasses information generated and collected through social media platforms such as Facebook, Instagram, and Twitter. Surprisingly, the reviewed literature does not show studies that use social media data exclusively. However, one study by Karthik et al. (2020) integrates both social media and non-social media data sources, including platforms such as Facebook and Google, as well as directories and services like Board, info@tds, Justdial, and Sulekha.

To date, social media data has been widely utilised in various domains, including the prediction of future sales and sentiment analysis (Asur & Huberman, 2010; Balfagih, 2016; Kanavos et al., 2023). For instance, a study by Asur & Huberman (2010) used Twitter data to forecast box-office revenues for movies based on tweet rates, which is an example of unstructured textual data. Similarly, a study by Balfagih (2016) imported data from his Facebook friend list to identify potential employees for a marketing team, which presents semi-structured user records. More recently, a study by Kanavos et al. (2023) analysed Twitter users' posts to detect spam content, which is another instance of unstructured textual data. Overall, social media data in other

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

domains tends to appear as unstructured or semi-structured data in the form of textual or user records.

However, using social media data presents several challenges. First, the biased population spectrum is observed across different social media platforms. One review paper clearly mentioned the differences in users' demographics when comparing Twitter, Facebook, Sina Weibo, WeChat, and Instagram (Huang et al., 2022). Thus, choosing the right social media platform as the data source is crucial, as it may impact the performance of lead scoring models. Second, the same review paper also stated the challenge of investigating multilingual data (Huang et al., 2022). In real-world scenarios, social media data often face challenges in contextual interpretation. However, to the extent of current knowledge, the data employed in sixteen lead scoring-related studies focus on only one language.

2.5 The Underexplored Potential of Social Media Data

Table 2 categorises the lead scoring-related papers based on data type (social media or non-social media) and source type (public or private). It distinguishes between these categories with counts for each combination.

Table 2. Dataset Allocation Table.

Data Type	Source Type	Count
Non-social media	Public	2
	Private	11
Social media and Non-social media	Public	0
	Private	1

As indicated in Table 2, fifteen out of sixteen (93.75%) studies in this area rely on non-social media data, and only one (6.25%) previous work by Karthik et al. (2020) utilised a mix of social media and non-social media data. Therefore, to the extent of current knowledge, there is a notable lack of research on lead scoring models utilising social media data. This gap in the literature suggests that current models might be missing out valuable insights derived from social media interactions and user behaviours. There is a need for more comprehensive studies to explore how incorporating social media data can enhance the accuracy and effectiveness of lead scoring models. Addressing this gap is essential for advancing the field, as integrating social media data could improve model performance and offer a more comprehensive understanding of lead-scoring dynamics.

In addition, only two out of fourteen (14.29%) datasets are publicly available. This gap highlights the challenge that researchers face due to the lack of accessible public datasets for experimental purposes in the context of lead scoring models. Moreover, this limitation complicates the assessment of a model's generalisation as it restricts the ability to test and validate models across diverse datasets.

3. Machine learning algorithms

ML is a subset of Artificial Intelligence that encompasses a vast array of algorithms and methodologies which are designed to enable computers to learn from data and make predictions or decisions (Helm et al., 2020). ML algorithms can be classified into two (2) categories: traditional ML algorithms and advanced ML algorithms. Traditional ML algorithms are simpler, more interpretable, and rely on feature engineering, including Logistic Regression (LR) (Fernandes et al., 2021), Naive Bayes (NB) (Berrar, 2018), Support Vector Machine (SVM) (Pisner & Schnyer, 2020), Decision Tree (DT) (Song & Lu, 2015), Ensemble Learning methods (Mienye & Sun, 2022) like Random Forest (RF) (A. B. Shaik & Srinivasan, 2019) and Gradient Boosting (GB) (Bentéjac et al., 2021). In contrast, advanced ML algorithms, such as Neural Networks (NN) (Bouwmans et al., 2019) and Deep Learning (DL) (Alzubaidi et al., 2021), are more complex, less interpretable, and capable of automatic feature learning but require significant computational resources (Thompson et al., 2021).

3.1 Traditional Machine Learning Algorithms

LR is specifically used for binary classification problems, where the outcome is categorical with two possible values (Arista, 2022; Fernandes et al., 2021; Zabor et al., 2022). NB is a classification algorithm that computes the probability of an event A given evidence B based on Bayes' Theorem (Berrar, 2018; Chen et al., 2021; Ismail et al., 2020). For example, calculating the probability of a lead being converted given their behaviour or activity data. SVM strives to identify the optimal hyperplane that maximises the margin between different classes, making it robust to outliers and effective for high-dimensional or non-linear data classification (Kim et al., 2020; Pisner & Schnyer, 2020; Zhu et al., 2023). DT is a highly efficient strategy for data mining, which organises a population into branch-like segments that form an inverted tree structure with a root node, internal nodes, and leaf nodes, thereby establishing a robust classification system (Aldino & Sulistiani, 2020; Lee et al., 2022; Song & Lu, 2015). For instance, DT is effective for organising leads into high- or low-potential groups by using a tree structure. Each branch of the tree represents a decision rule based on lead attributes such as engagement, demographics, and past behaviour. The root node might split leads by engagement level, with branches further dividing them by job title. This hierarchy helps categorise leads into segments, making it easy to identify and prioritise them.

Ensemble Learning is the synergy of two or more ML algorithms, which combines their strengths to achieve superior performance beyond what each algorithm could achieve alone (Mienye & Sun, 2022; Mohammed & Kora, 2023; Zhao et al., 2020). RF is an ensemble method that constructs and intertwines multiple decision trees. RF is capable of handling large datasets and operates without the necessity for cross-validation or data preprocessing techniques such as data rescaling and transformation (Bhardwaj et al., 2022; Mohd Ali et al., 2022; A. B. Shaik & Srinivasan, 2019). GB is another ensemble ML algorithm, which combines several weak prediction models, often simple DTs, to form a stronger model, sequentially training each new model to correct the errors of previous models (Bentéjac et al., 2021; Nie et al., 2021; Sahin, 2020).

3.2 Advanced Machine Learning Algorithms

While traditional ML algorithms are effective for many tasks, advanced ML algorithms have emerged to address complex problems. NN was inspired by the structure and function of the human brain, which consists of interconnected nodes referred to

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

as neurons, and eventually arranged into layers (Bouwmans et al., 2019; Thakur & Konde, 2021; Yang & Wang, 2020). Each neuron receives input, processes it through hidden layers using weights, and produces an output. DL employs a multilayered NN to enable the simultaneous process of learning and classification, effectively addressing the issues of incorrect predictions caused by biased feature selection (Alzubaidi et al., 2021). DL architectures come in a variety of forms, each tailored to specific tasks, including DNN (Srinidhi et al., 2021), FNN (Puravankara & Narendra Babu, 2020), RNN (Shiri et al., 2023), Long Short-Term Memory (LSTM) (DiPietro & Hager, 2020), and Kolmogorov–Arnold Networks (KAN) (Liu et al., 2024).

FNN is a simple and straightforward NN that contains one or multiple hidden layers with direct connections between units and information travelling only in a forward direction (Abdel-Nasser Sharkawy, 2020; Markowska-Kaczmar & Kosturek, 2021; Wang et al., 2023). RNN can effectively handle sequential relationships due to its backward linkages from the output to the input and hidden layers (Kane & Hussain, 2023; Shiri et al., 2023), which has been applied in various civil engineering domains (Mers et al., 2022). LSTM is a subset of RNN designed to address the vanishing gradient problem with the usage of memory units and gating mechanisms to manage long-term dependencies (Nosouhian et al., 2021; Shiri et al., 2023), making it well-suited for healthcare, energy and natural language processing domains (Al-Selwi et al., 2024). DNN is a sophisticated, multilayered NN composed of several hidden layers, including input and output layers (Sarker, 2021; Vijayakumar et al., 2021). It is capable of learning complex patterns from input data without requiring manual feature engineering and has been widely used for time-series prediction and classification tasks (How et al., 2020; Sarker, 2021; Vijayakumar et al., 2021). KAN, inspired by the Kolmogorov-Arnold representation theorem, enhances model flexibility by replacing the fixed linear weights in traditional NN with learnable activation functions, which excel in handling time series forecasting (Hou et al., 2024; Liu et al., 2024).

While there is no direct comparison of the computational costs associated with the relevant studies due to the lack of such data in the cited studies, advanced ML algorithms are often believed to have a longer training time and higher computational costs compared to traditional ML algorithms (Ali et al., 2024).

4. Feature Engineering Techniques

Feature engineering involves preparing datasets to optimise a model's performance by transforming, extracting, and selecting relevant data features (Kumar et al., 2023). Feature transformation refers to the process of converting existing features into a more suitable form for modelling, which includes techniques like normalisation, scaling, and encoding (Das & Spanos, 2022; Zhuang et al., 2020). Feature extraction can be defined as a process of deriving new features from existing ones, utilising methods such as principal component analysis (PCA), latent semantic indexing (LSI), and clustering methods (Suhaidi et al., 2021). For instance, PCA can help to identify the most important features that have the largest covariance with lead outcomes (C. Zhang et al., 2018), LSI works well with textual lead data in a semantic space to capture relationships between terms and documents (Zelikovitz & Hirsh, 2001), and clustering methods group leads with similar characteristics, allowing for more effective segmentation and prioritisation based on lead potential (Bharti & Singh, 2015). Feature selection is the process of reducing data dimensionality by identifying and selecting relevant features in order to minimise overfitting, increase accuracy, and reduce training time (Kumar et al., 2023; T. Shaik et al., 2022).

5. Discussions and analysis

5.1 The Debate on Lead Scoring: Is it Truly Classification?

Table 3 categorises the lead scoring-related papers based on the ML task, distinguishing between classification and regression.

Table 3. ML Task Allocation: Classification or Regression.

ML Task	Class	Count
Classification	Binary	16
	Non-Binary	0
Regression	Continuous Scores	0

Examining Table 3, it is apparent that all studies focus on predicting binary classes using classification techniques. Despite the term lead "scoring", these models often produce binary classification outputs rather than assigning continuous scores (Ayaz, 2023; Benhaddou & Leray, 2017; Bhatta, 2022; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Espadinha-Cruz et al., 2021; Etminan, 2021; Giorcelli, 2019; Jadli et al., 2022; Karthik et al., 2020; Nygård & Mezei, 2020; Pereira, 2021; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Swelsen, 2019; Yim et al., 2024). This observed pattern offers a descriptive analysis of current methodological practices, where the reviewed literature consistently uses the term "scoring" to describe binary classification frameworks. However, in its typical mathematical usage, the term "scoring" generally implies continuous measurement.

In fact, lead scoring is typically a classification task to simplify the decision-making process. A possible explanation for using a binary outcome, such as "convert" or "not convert", is that it provides a clear and actionable decision point to prioritise potential leads more effectively. However, reducing lead scoring to a binary outcome may result in the loss of valuable information about how promising each lead truly is. For instance, with continuous scoring, the lead who frequently visits the pricing page could receive a higher score of 85 out of 100, reflecting a stronger likelihood of conversion compared to the lead with fewer interactions, who might receive a score of 60. This more granular approach enables decision-makers to prioritise leads more effectively. Sales teams could allocate more resources to the higher-scoring leads, potentially increasing conversion rates and improving overall sales efficiency. Continuous scoring, therefore, provides more nuanced insights into the likelihood of conversion, leading to more refined decision-making and optimised resource allocation. This discussion highlights the trade-off between simplicity and granularity in lead scoring approaches.

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

5.2 Comparative Analysis of "X Online Education" Dataset

By far, the most widely used dataset for lead scoring models is the publicly available "X Online Education" dataset, which was employed in four different research papers (Jadli et al., 2022; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Yim et al., 2024). This dataset[1] is available on Kaggle, an online community that allows users to find and publish datasets. The dataset pertains to an education company, known as X Education, which offers online courses tailored for industry professionals. The company collects data through multiple channels, including forms filled out by website visitors and past referrals. The collected data is used to classify individuals into two categories: converted (those who enrol in courses) or not converted (those who do not enrol) after being approached by the sales team.

Even when using the same dataset, the findings differ across the four papers. Table 4 presents that two unique works by Jadli et al. (2022) and M. Sharma (2022) reported that RF demonstrated the highest accuracy in predicting lead scoring

Table 4. Comparison of the Accuracy Score between four papers that used the same "X Online Education" dataset. The highest results achieved by each algorithm will be bolded, while the highest results achieved by each study will be marked with an asterisk (*) at the end.

Algorithm	Yim et al. (2024)	M. Sharma (2022)	Jadli et al. (2022)	Puravankara & Narendra Babu (2020)
AdaBoost	-	-	-	**0.929**
Bagging	**0.9216**	-	-	-
Boosting	**0.9116**	-	-	-
CatBoost	-	**0.8502**	-	-
DT	-	0.7911	**0.9147**	-
Feedforward NN	-	-	-	**0.94***
GB	-	-	-	**0.933**
KNN	**0.9197**	-	0.8004	-
LR	0.9207	0.8326	0.8989	**0.923**
NB	**0.9142**	-	0.8018	-
RF	0.9222	0.8528*	**0.9302***	0.925
Stacking	**0.9233***	-	-	-
SVM	**0.9215**	-	0.7322	-
XGBoost	-	0.8517	-	**0.935**

Table 5. Comparison of Feature Engineering Techniques between four papers that used the same "X Online Education" dataset.

Feature Engineering Techniques	Yim et al. (2024)	M. Sharma (2022)	Jadli et al. (2022)	Puravankara & Narendra Babu (2020)
Check duplicates	✓	✓	✓	✓
Drop ID columns (prospect id, lead number)	✓			✓
Replace unselected value to NaN			✓	
Drop columns with null	✓ (drop if more than 40%)	✓ (drop if more than 3,000)	✓ (drop if more than 70%)	✓ (drop if more than 45%)
Replace null value	✓ (data imputation, mode)	✓ (data imputation, mode)	✓ (mode, mean, median)	✓ (mode)
Check correlations	✓			✓
Drop columns with same values		✓ (drop if more than 90%)		
Check data skewness		✓	✓	
Check outliers	✓	✓	✓	✓
Remove outliers	✓	✓		
Transform categorical to numeric	✓ (dummy variable creation)	✓ (one hot encoding)	✓ (one hot encoding)	✓ (dummy variable creation)
Scaling	✓ (standard)	✓ (min max)		✓
Normlisation	✓	✓		
K-fold	✓		✓	
Split	80/20	70/30	70/30	80/20
Feature Elimination	✓ (Recursive Feature Elimination)			

[1] X Online Education dataset link: https://www.kaggle.com/datasets/amolbhone/lead-score-case-study?select=lead+scoring+case+study.pdf

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

outcomes. Meanwhile, a study by Yim et al. (2024) diverged in the finding, identifying that Stacking, one of the Ensemble Learning algorithms, was the most effective. Moreover, a study by Puravankara and Narendra Babu (2020) reveals that FNN performs the best in predicting lead conversion. Overall, the best model, according to performance metrics, was produced by Puravankara and Narendra Babu (2020) with the usage of the FNN algorithm.

The variation in the performance findings highlights the influence of different factors on model performance, such as data preprocessing techniques, specific implementation and fine-tuning of the algorithms. Consistent data preprocessing techniques are crucial, as they help ensure the quality and reliability of the data, thereby reducing bias and enhancing the model's generalisability across different datasets. Looking at Table 5, the work of Puravankara and Narendra Babu (2020) has fewer data preprocessing steps when compared to other works. This study did not replace unselected values with NaN values, drop columns with identical values, check for data skewness, remove outliers, apply normalisation, perform k-fold cross-validation, or conduct feature elimination. It might imply that too many data preprocessing techniques are unnecessary for handling a lead scoring dataset.

For comparison purposes, LR and RF are selected since all four studies implemented these algorithms. Among the studies using both LR and RF algorithms, the work of Puravankara and Narendra Babu (2020) achieved the highest performance, followed by the work of Yim et al. (2024). The common data preprocessing techniques used by these two studies but not by the other two include dropping ID columns, removing columns that have 40% to 45% null values, using mode to replace null value, checking correlations, not checking for data skewness, using dummy variable creation to transform categorical values into numeric values, and splitting the data into training and test sets with a ratio of 80:20. These techniques are likely contributed to their superior results compared to other studies that did not use these methods.

5.2 Current Trend on Predictive Lead Scoring Models

This subsection focuses on trends related to the application of algorithms and performance metrics. It will first discuss the trend between traditional and advanced ML algorithms. Next, the subsection will delve into each algorithm, including both traditional and advanced ones.

Figure 3 illustrates the trends of ML types, including traditional and advanced ML algorithms, from 2013 to 2024. Based on Figure 3, it is evident that traditional ML algorithms play a significant role in the development of lead scoring models and are the preferred choice for researchers. Despite the effectiveness of advanced ML algorithms, traditional ML algorithms remain popular. A great deal of previous research works (Ayaz, 2023; Balfagih, 2016; Bhatta, 2022; Choudhury & Nur, 2019; D'Haen & Van Den Poel, 2013; Espadinha-Cruz et al., 2021; Etminan, 2021; Jadli et al., 2022; Nygård & Mezei, 2020; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Yim et al., 2024) have focused on implementing traditional ML algorithms. This preference may be attributed to their ease of implementation, interpretability, and lower computational requirements (G. Zhang, 2023).

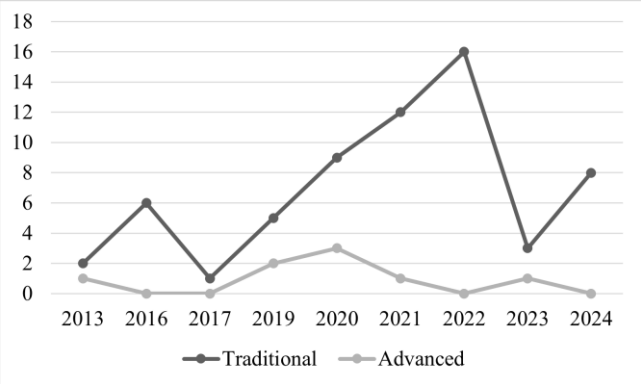

Figure 3. Trends of ML Types in Lead Scoring Models.

However, traditional ML algorithms are not necessarily the best-performing ones, despite their popularity. Up to now, several studies have indicated that advanced ML algorithms achieved superior performance compared to traditional ML algorithms in building predictive lead scoring models (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Puravankara & Narendra Babu, 2020). Indeed, advanced ML algorithms excel in handling large datasets (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Puravankara & Narendra Babu, 2020), which involve a high degree of complexity, manage a high volume of variables, and capture complex non-linear relationships between features and targets. However, their "black box" nature poses challenges in understanding, explaining, and interpreting the models, which may contribute to their lower popularity (Koeppe et al., 2022). This mixed performance, along with the higher complexity and computational requirements of advanced algorithms, contributes to the ongoing preference for traditional ML methods.

In some cases, advanced ML algorithms are not always the best-performing algorithms. The work by D'Haen and Van Den Poel (2013) concluded that DT outperformed NN and LR by initiating a feedback loop to optimise and stabilise the model. Nygård and Mezei (2020) found that RF outperformed LR, DT, and NN in terms of overall performance. Although NN had higher accuracy and specificity, RF had higher AUC and sensitivity. Furthermore, Espadinha-Cruz et al. (2021) stated that Ensemble Learning was the best-performing algorithm. Thus, the three individual studies make it difficult to conclusively state that advanced ML algorithms are more effective overall.

The findings from the datasets used in these studies indicate that NN tend to outperform in scenarios involving larger datasets (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Puravankara & Narendra Babu, 2020). Conversely, studies with smaller datasets often find traditional ML algorithms to be more effective (D'Haen & Van Den Poel, 2013; Nygård & Mezei, 2020). In contrast, the study by Espadinha-Cruz et al. (2021) demonstrates that an ensemble approach, combining regression with PCA and NN using varying numbers of neurons in the hidden layer, outperforms a standalone NN.

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Figure 4 provides a comparative synthesis, delineating the algorithms used in lead scoring models across the sixteen studies. The examined papers underscore a diverse range of ML algorithms, spanning from traditional ML to Ensemble Learning and advanced ML algorithms.

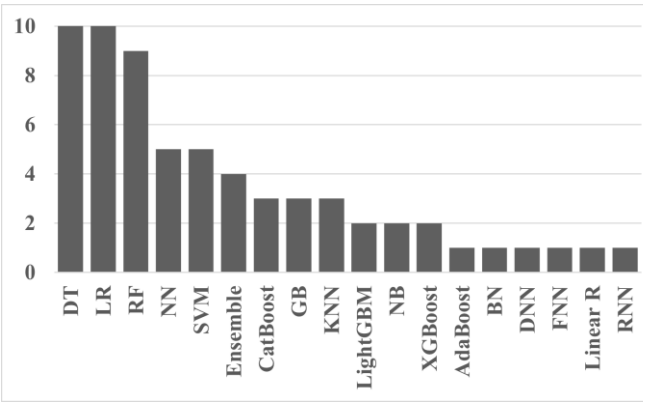

Figure 4. Trends of ML Algorithms Used in Lead Scoring Models.

Figure 4 proves that DT, LR, and RF are among the most frequently implemented algorithms in lead scoring models. Perhaps, the popularity of LR can likely be attributed to its effectiveness in solving linear relationships between features and targets, which aligns well with the binary outputs of lead scoring models. On the other hand, DT is prone to overfitting but is advantageous for handling non-linear relationships between features. RF, an ensemble method that combines multiple DTs, tends to offer improved performance over DT.

5.3 Performance Metrics in Existing Lead Scoring Models

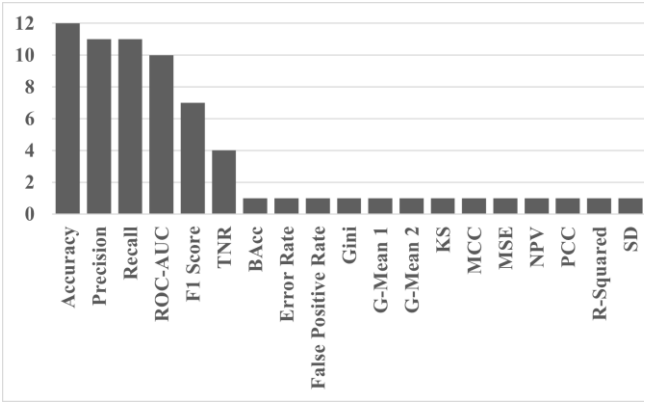

Figure 5. Trends of Performance Metrics Used in Lead Scoring Models.

Figure 5 shows that, among 19 performance metrics utilised across sixteen studies, accuracy is the most frequently used metric (17.65%). Since the studies did not employ a common performance metric for model evaluation, this paper will focus on the most commonly used metric, which is accuracy. The highest accuracy score reported is 99.41%, which is achieved by the work of Choudhury & Nur (2019). However, due to the variation in datasets used across the studies, direct performance comparisons should only be made when the same dataset is applied.

5.4 Feature Engineering Techniques in Existing Lead Scoring Models

Figure 6 compares the different approaches employed by sixteen studies to handle feature engineering.

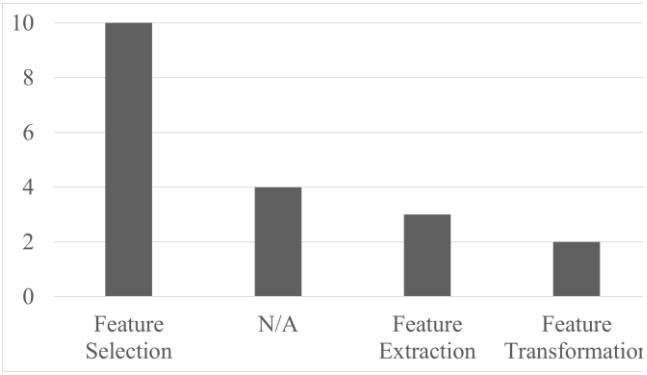

Figure 6. Trends of Feature Engineering Techniques Used in Lead Scoring Models.

From Figure 6, twelve out of sixteen (75%) studies utilised one or more feature engineering techniques. Among these, feature selection emerged as the most popular method, as it effectively reduces the dataset's dimensionality and avoids overfitting issues by choosing relevant features.

6. Challenges and future work

Despite the promising results reported in the studies, several challenges or limitations were noted and are classified in Table 6, which will be discussed in the following subsections.

Table 6. Challenges/Limitations faced by Related Work.

Challenges/Limitations Faced	Related Work
Dataset issues	(Ayaz, 2023; Benhaddou & Leray, 2017; Bhatta, 2022; Espadinha-Cruz et al., 2021; Etminan, 2021; Giorcelli, 2019; Pereira, 2021; Puravankara & Narendra Babu, 2020; Swelsen, 2019; Yim et al., 2024)
Hyperparameter Tuning issues	(Benhaddou & Leray, 2017; Bhatta, 2022; Choudhury & Nur, 2019; Giorcelli, 2019; Jadli et al., 2022; Karthik et al., 2020; Nygård & Mezei, 2020; Swelsen, 2019; Yim et al., 2024)
Performance issues	(Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; D'Haen & Van Den Poel, 2013; Jadli et al., 2022; Karthik et al., 2020; Pereira, 2021)

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

6.1 Dataset Issues

One recurring issue is dataset issues, including imbalanced data (Benhaddou & Leray, 2017; Etminan, 2021; Swelsen, 2019; Yim et al., 2024), limited size (Benhaddou & Leray, 2017; Etminan, 2021; Giorcelli, 2019), and low quality of data (Pereira, 2021), which may eventually lead to biased predictions (Ayaz, 2023; Yim et al., 2024). Other reported dataset issues include a lack of relationships between datasets (Puravankara & Narendra Babu, 2020), the use of real customer data instead of lead data (Swelsen, 2019), and insufficient data visibility (Bhatta, 2022).

A prior study by Pereira (2021) struggled to achieve high model performance due to low-quality data with only 56% accuracy, 71% sensitivity, 49% specificity, and 42% precision. The low performance of the model was attributed to poor data quality, characterised by a significant proportion of missing values, incomplete data, and inappropriate data formats. Similarly, another study by Ayaz (2023) emphasised that the effectiveness of lead scoring models heavily relies on the quality of the data used. The possible reasons for these dataset issues are problems during the data collection phase, such as a limited data collection period, inadequate data validation processes, and the absence of relevant data. Such issues will lead to imbalances, a limited dataset size, and low data quality, collectively contributing to biased or unreliable model predictions. These studies underscore the critical importance of high-quality data for the success of building a lead scoring model.

Suggestions for future work include thoroughly planning the data collection phase to avoid insufficient and incomplete data. Besides that, Generative Adversarial Networks, a DL model that generates synthetic data mimicking real data, might help solve dataset issues (Akkem et al., 2024). GAN consists of two components: a generator, which produces synthetic data by sampling from a random prior noise distribution, and a discriminator, which evaluates whether the data produced by the generator is real or fake (Navidan et al., 2021). Thus, GAN can address issues such as data imbalance, limited dataset size and low data quality by generating more high-quality lead examples.

6.2 Hyperparameter Tuning Issues

Hyperparameter tuning issues pose significant challenges or limitations. Several studies have mentioned using default parameter values (Benhaddou & Leray, 2017; Choudhury & Nur, 2019; Giorcelli, 2019; Jadli et al., 2022; Karthik et al., 2020; Nygård & Mezei, 2020; Swelsen, 2019; Yim et al., 2024), indicating that hyperparameter tuning was not implemented. As a result, the reported outcomes may not represent the model's optimal performance. A study by Bhatta (2022) mentioned the trade-off between comprehensive hyperparameter optimisation and limited computing resources, which led to the usage of a smaller grid search. This issue can lead to missed opportunities for finding the optimal parameter values, resulting in models that are less effective than they could be with more extensive tuning.

Future research should focus on using advanced hyperparameter optimisation techniques to maximise model performance. Additionally, there is a need to develop more efficient and automated methods for hyperparameter tuning to save time and computational resources. For instance, the Hyperband algorithm, combined with the Bayesian optimisation method, which leverages the strengths of both standalone methods, often outperforms other approaches such as Random Search, as well as each method, and demonstrates a greater margin of improvement, specifically in more complex optimisation problems (Wang Blippar et al., 2018). This approach would enable the optimisation process of lead scoring models to allocate resources to the most promising hyperparameters dynamically. By leveraging the Hyperband algorithm with the Bayesian optimisation method, the search can be refined based on past trials, which significantly speeds up the tuning process and reduces computational costs.

6.3 Performance Issues

Many studies have identified performance issues as one of the challenges or limitations faced. These issues include not addressing the possibility of overfitting in high-performance models (Choudhury & Nur, 2019; Jadli et al., 2022; Karthik et al., 2020), models not being fully tested in real-time environments (Chaudhuri et al., 2021; D'Haen & Van Den Poel, 2013; Pereira, 2021), and the trade-off between accuracy and interpretability (Ayaz, 2023).

The possibility of overfitting in high-performance models indicates that the model may be too closely tailored to the training data, thereby leading to poor performance on new, unseen data. For instance, if a lead scoring model is overfitted, it may misclassify potentially high-value leads as low-value simply because their characteristics do not match the training data. Conversely, it might incorrectly classify low-value leads as high-value, leading to wasted resources on leads that are unlikely to convert. This undermines the generalisability of the model, making it less reliable for practical applications. The lack of testing the models in real-time environments raises questions about the practical applicability of these models, as a model may perform well in a controlled setting but not in a dynamic, real-time situation. While complex models using advanced ML algorithms often yield higher accuracy, they do so at the cost of interpretability. Models that lack interpretability may not earn users' trust as they cannot discern the factors driving the model's outputs.

To date, DL algorithms have yet to be widely integrated into lead scoring models. In fact, despite being ten times slower than Multi-layer Perceptrons, KAN, a newly introduced NN algorithm, offers better accuracy for science-related tasks, making it a promising alternative for future work where accuracy may outweigh computational speed (Liu et al., 2024). Aside from that, Hyperbolic Deep Learning (HNN), which employs hyperbolic space rather than traditional Euclidean space, warrants further analysis. The work by Ganea et al. (2018) indicates that HNN often outperforms its Euclidean variants in most settings. HNN has the potential to significantly improve model accuracy, particularly in tasks involving complex hierarchical or relational data (Ganea et al., 2018), as hyperbolic space is well-suited for representing tree-like structures (Chami et al., 2020). Furthermore, HNN's ability to handle large and complex datasets could make it a powerful tool

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

for improving accuracy in lead scoring models, especially as data becomes more abundant and intricate in future work.

7. Conclusion

This review paper briefly introduces the concepts of lead scoring, ML algorithms, and feature engineering techniques. It also identifies common characteristics of datasets used in lead scoring models, such as reliance on non-social media data and the lack of accessible public datasets. Besides that, this review paper includes a classification table to summarise the lead scoring-related papers based on the ML task, either classification or regression. Apart from that, a comprehensive table provides an overview of different studies utilising the same "X Online Education" dataset. Despite using the same dataset, the results varied due to differences in data preprocessing and feature selection techniques employed by the authors.

This review paper summarises key findings on trends in lead scoring models, highlighting the popularity of traditional ML algorithms, particularly DT, LR, and RF algorithms. Furthermore, this review paper justifies the trends in feature engineering techniques used in lead scoring models.

The paper continues to suggest future directions for building lead scoring models, such as planning the data collection phase thoroughly, using advanced hyperparameter optimisation techniques, and implementing DL algorithms. Nevertheless, more research is required in order to conclude whether the suggested solutions are able to enhance the performance of lead scoring models.

In brief, this review paper plays a pivotal role in advancing the field of lead scoring by offering valuable insights for both academics and practitioners seeking to refine their strategies. By optimising lead scoring, businesses can streamline their sales processes, prioritise high-potential leads, and ultimately enhance overall efficiency and profitability, driving economic growth and job creation. As the landscape of data and technology continues to evolve, ongoing research is essential to refine these strategies further and unlock even greater potential for business success.

8. Acknowledgement

This research was supported by Tunku Abdul Rahman University of Management and Technology (TAR UMT) through the grant UC/I/G2022-00098.

9. References

Abdel-Nasser Sharkawy. (2020). Principle of Neural Network and Its Main Types: Review. *Journal of Advances in Applied & Computational Mathematics*, *7*, 8–19. https://doi.org/10.15377/2409-5761.2020.07.2

Akkem, Y., Biswas, S. K., & Varanasi, A. (2024). A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. *Engineering Applications of Artificial Intelligence*, *131*, 107881. https://doi.org/10.1016/J.ENGAPPAI.2024.107881

Aldino, A. A., & Sulistiani, H. (2020). Decision Tree C4.5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia). *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, *7*(1), 40–50. https://doi.org/10.21107/EDUTIC.V7I1.8849

Ali, A., Jayaraman, R., Azar, E., & Maalouf, M. (2024). A comparative analysis of machine learning and statistical methods for evaluating building performance: A systematic review and future benchmarking framework. *Building and Environment*, *252*, 111268. https://doi.org/https://doi.org/10.1016/j.buildenv.2024.111268

Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., Muneer, A., Sumiea, E. H., Alqushaibi, A., & Ragab, M. G. (2024). RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. *Journal of King Saud University - Computer and Information Sciences*, *36*(5), 102068. https://doi.org/https://doi.org/10.1016/j.jksuci.2024.102068

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Arista, A. (2022). Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19. *Jurnal Dan Penelitian Teknik Informatika*, *7*(1). https://doi.org/10.33395/sinkron.v7i1.11243

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*. https://doi.org/10.1109/wi-iat.2010.63

Autopilot. (2016, January 12). *Lead Scoring is Broken. Here's What to Do Instead.* https://medium.com/marketing-on-autopilot/lead-scoring-is-broken-here-s-what-to-do-instead-194a0696b8a3

Ayaz, S. B. (2023). *Lead Scoring with Machine Learning*.

Balfagih, A. (2016). *Direct Selling Business Lead Prediction by Social Media Data Mining*. https://DalSpace.library.dal.ca//handle/10222/71412

Benhaddou, Y., & Leray, P. (2017). Customer relationship management and small data - Application of Bayesian network elicitation techniques for building a lead scoring model. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, *2017-October*, 251–255. https://doi.org/10.1109/AICCSA.2017.51

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Berrar, D. (2018). *Bayes' Theorem and Naive Bayes Classifier*. https://doi.org/10.1016/B978-0-12-809633-8.20473-1

Bhardwaj, P., Tiwari, P., Olejar, K., Parr, W., & Kulasiri, D. (2022). A machine learning application in wine quality prediction. *Machine Learning with Applications*, *8*, 100261. https://doi.org/https://doi.org/10.1016/j.mlwa.2022.100261

Bharti, K. K., & Singh, P. K. (2015). Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. *Expert Systems with Applications*, *42*(6), 3105–3114. https://doi.org/https://doi.org/10.1016/j.eswa.2014.11.038

Bhatta, I. (2022). *Optimizing Marketing Channel Attribution for B2B and B2C with Machine Learning Based Lead Scoring Model*.

Bouwmans, T., Javed, S., Sultana, M., & Jung, S. K. (2019). Deep neural network concepts for background subtraction:A systematic review and comparative evaluation. *Neural Networks*, *117*, 8–66. https://doi.org/10.1016/J.NEUNET.2019.04.024

Ceccatelli, J. (2023). *The impact of lead scoring on CRM and the sales process*. https://www.repository.utl.pt/handle/10400.5/28214

Chami, I., Gu, A., Chatziafratis, V., & Ré, C. (2020). From Trees to Continuous Embeddings and Back: Hyperbolic Hierarchical Clustering. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 15065–15076). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/ac10ec1ace51b2d973cd87973a98d3ab-Paper.pdf

Chaudhuri, N., Gupta, G., Vamsi, V., & Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems*, *149*, 113622. https://doi.org/10.1016/J.DSS.2021.113622

Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *Eurasip Journal on Advances in Signal Processing*, *2021*(1), 1–12. https://doi.org/10.1186/S13634-021-00742-6/TABLES/5

Choudhury, A. M., & Nur, K. (2019). A Machine Learning Approach to Identify Potential Customer Based on Purchase Behavior. *2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)*, 242–247. https://doi.org/10.1109/ICREST.2019.8644458

Das, H. P., & Spanos, C. J. (2022). Improved dequantization and normalization methods for tabular data pre-processing in smart buildings. *BuildSys 2022 - Proceedings of the 2022 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 168–177. https://doi.org/10.1145/3563357.3564072

De Haan, E., & Menichelli, E. (2019). *The Incremental Value of Unstructured Data in Predicting Customer Churn*.

D'Haen, J., & Van Den Poel, D. (2013). Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework. *Industrial Marketing Management*, *42*(4), 544–551. https://doi.org/10.1016/J.INDMARMAN.2013.03.006

DiPietro, R., & Hager, G. D. (2020). Chapter 21 - Deep learning: RNNs and LSTM. In S. K. Zhou, D. Rueckert, & G. Fichtinger (Eds.), *Handbook of Medical Image Computing and Computer Assisted Intervention* (pp. 503–519). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-816176-0.00026-0

Duncan, B., & Elkan, C. (2015). Probabilistic modeling of a sales funnel to prioritize leads. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *2015-August*, 1751–1758. https://doi.org/10.1145/2783258.2788578

Espadinha-Cruz, P., Fernandes, A., & Grilo, A. (2021). Lead management optimization using data mining: A case in the telecommunications sector. *Computers & Industrial Engineering*, *154*, 107122. https://doi.org/10.1016/J.CIE.2021.107122

Etminan, A. (2021). *Prediction of Lead Conversion With Imbalanced Data : A method based on Predictive Lead Scoring*. https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-176433

Fahad, K. M. W. (2017). *A study on business development strategies of LEADS Impacto Ltd*. http://dspace.bracu.ac.bd:8080/xmlui/handle/10361/8694

Fernandes, A. A. T., Filho, D. B. F., da Rocha, E. C., & da Silva Nascimento, W. (2021). Read this paper if you want to learn logistic regression. *Revista de Sociologia e Política*, *28*(74), 006. https://doi.org/10.1590/1678-987320287406EN

Ganea, O.-E., Bécigneul, G., & Hofmann, T. (2018). Hyperbolic Neural Networks. *Advances in Neural Information Processing Systems*, *31*.

Giorcelli, G. (2019). *Variable-sized input, character-level recurrent neural networks in lead generation: predicting close rates from raw user inputs*. https://arxiv.org/abs/1901.05115v1

Goh, K. H., Wang, L., Yeow, A. Y. K., Poh, H., Li, K., Yeow, J. J. L., & Tan, G. Y. H. (2021). Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nature Communications*, *12*(1), 711. https://doi.org/10.1038/s41467-021-20910-4

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Current Reviews in Musculoskeletal Medicine*, *13*(1), 69–76. https://doi.org/10.1007/s12178-020-09600-8

Hou, Y., zhang, D., Wu, J., & Feng, X. (2024). *A Comprehensive Survey on Kolmogorov Arnold Networks (KAN)*. https://arxiv.org/abs/2407.11075v3

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

How, D. N. T., Hannan, M. A., Lipu, M. S. H., Sahari, K. S. M., Ker, P. J., & Muttaqi, K. M. (2020). State-of-Charge Estimation of Li-Ion Battery in Electric Vehicles: A Deep Neural Network Approach. *IEEE Transactions on Industry Applications*, *56*(5), 5565–5574. https://doi.org/10.1109/TIA.2020.3004294

Huang, X., Wang, S., Zhang, M., Hu, T., Hohl, A., She, B., Gong, X., Li, J., Liu, X., Gruebner, O., Liu, R., Li, X., Liu, Z., Ye, X., & Li, Z. (2022). Social media mining under the COVID-19 context: Progress, challenges, and opportunities. *International Journal of Applied Earth Observation and Geoinformation*, *113*, 102967. https://doi.org/https://doi.org/10.1016/j.jag.2022.102967

Ismail, M., Hassan, N., & Saleh Bafjaish, S. (2020). Comparative Analysis of Naive Bayesian Techniques in Health-Related For Classification Task. *Journal of Soft Computing and Data Mining*, *1*(2), 1–10. https://doi.org/10.30880/jscdm.2020.01.02.001

Jadli, A., Hamim, M., Hain, M., & Hasbaoui, A. (2022). *TOWARD A SMART LEAD SCORING SYSTEM USING MACHINE LEARNING*. https://doi.org/10.21817/indjcse/2022/v13i2/221302098

Kanavos, A., Antonopoulos, N., Karamitsos, I., & Mylonas, P. (2023). A Comparative Analysis of Tweet Analysis Algorithms Using Natural Language Processing and Machine Learning Models. *2023 18th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2023*. https://doi.org/10.1109/SMAP59435.2023.10255184

Kane, A., & Hussain, A. (2023). Artificial Neural Networks: An Overview. *Mesopotamian Journal of Computer Science*, *2023*(5), 124–133. https://doi.org/10.58496/MJCSC/2023/015

Karthik, P. V. V., Rao, P. G. P., & Narsimlu, M. (2020). Case Study: Lead Conversion of Digital Marketing Data Using Predictive Analytics. *Lecture Notes in Electrical Engineering*, *601*, 510–519. https://doi.org/10.1007/978-981-15-1420-3_54

Kim, D., Kang, S., & Cho, S. (2020). Expected margin–based pattern selection for support vector machines. *Expert Systems with Applications*, *139*, 112865. https://doi.org/10.1016/J.ESWA.2019.112865

Koeppe, A., Bamer, F., Selzer, M., Nestler, B., & Markert, B. (2022). Explainable Artificial Intelligence for Mechanics: Physics-Explaining Neural Networks for Constitutive Models. *Frontiers in Materials*, *8*, 824958. https://doi.org/10.3389/FMATS.2021.824958/BIBTEX

Kumar, K. P., Unal, A., Pillai, V. J., Murthy, H., & Niranjanamurthy, M. (2023). *Data engineering and data science: concepts and applications*.

Lee, C. S., Yeng, P., Cheang, S., & Moslehpour, M. (2022). *Predictive Analytics in Business Analytics: Decision Tree*.

Lindahl, E., Skolan, K., Datavetenskap, F., & Kommunikation, O. (2017). *A qualitative examination of lead scoring in B2B marketing automation, with a recommendation for its practice.* https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-213432

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2024). *KAN: Kolmogorov-Arnold Networks*. https://arxiv.org/abs/2404.19756v2

Markowska-Kaczmar, U., & Kosturek, M. (2021). Extreme learning machine versus classical feedforward network. *Neural Computing and Applications*, *33*(22), 15121–15144. https://doi.org/10.1007/s00521-021-06402-y

Mers, M., Yang, Z., Hsieh, Y.-A., & Tsai, Y. (James). (2022). Recurrent Neural Networks for Pavement Performance Forecasting: Review and Model Performance Comparison. *Transportation Research Record*, *2677*(1), 610–624. https://doi.org/10.1177/03611981221100521

Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, *10*, 99129–99149. https://doi.org/10.1109/ACCESS.2022.3207287

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/J.JKSUCI.2023.01.014

Mohd Ali, N. F., Mohd Sadullah, A. F., Abdul Majeed, A. P. P., Mohd Razman, M. A., & Musa, R. M. (2022). The identification of significant features towards travel mode choice and its prediction via optimised random forest classifier: An evaluation for active commuting behavior. *Journal of Transport & Health*, *25*, 101362. https://doi.org/10.1016/J.JTH.2022.101362

Morgado, A. V. (2018). The Value of Customer References to Potential Customers in Business Markets. *Journal of Creating Value*, *4*(1), 132–154. https://doi.org/10.1177/2394964318771799

Navidan, H., Moshiri, P. F., Nabati, M., Shahbazian, R., Ghorashi, S. A., Shah-Mansouri, V., & Windridge, D. (2021). Generative Adversarial Networks (GANs) in networking: A comprehensive survey & evaluation. *Computer Networks*, *194*, 108149. https://doi.org/https://doi.org/10.1016/j.comnet.2021.108149

Nie, P., Roccotelli, M., Fanti, M. P., Ming, Z., & Li, Z. (2021). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports*, *7*, 1246–1255. https://doi.org/10.1016/J.EGYR.2021.02.006

Niemi, E. (2022). *Sales Lead Utilization with Distribution Partners in B2B : Creating an Analytical Model for Sales Lead Assessment*. https://trepo.tuni.fi/handle/10024/136382

Nosouhian, S., Nosouhian, F., & Khoshouei, A. K. (2021). A Review of Recurrent Neural Network Architecture for Sequence Learning: Comparison between LSTM and GRU. *Preprints*. https://doi.org/10.20944/preprints202107.0252.v1

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Nygård, R., & Mezei, J. (2020). Automating Lead Scoring with Machine Learning: An Experimental Study. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-January*, 1439–1448. https://doi.org/10.24251/HICSS.2020.177

Pereira, R. M. M. (2021). *Building a predictive lead scoring model for contact prioritization : the case of HUUB*. https://repositorio.ucp.pt/handle/10400.14/34877

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. *Machine Learning: Methods and Applications to Brain Disorders*, 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Puravankara, R., & Narendra Babu, C. (2020). Lead Forecasting using LSTM based Deep Learning Architecture for Sentiment Analysis. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, 159–164. https://doi.org/10.1109/ICOIACT50329.2020.9332092

Rodrigues, J. D. da S. S. (2020). *Automated lead scoring system: a case study of a Portuguese startup*. Universidade do Porto (Portugal).

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, *2*(7), 1–17. https://doi.org/10.1007/S42452-020-3060-1/TABLES/1

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Shaik, A. B., & Srinivasan, S. (2019). A Brief Survey on Random Forest Ensembles in Classification Model. *Lecture Notes in Networks and Systems*, *56*, 253–260. https://doi.org/10.1007/978-981-13-2354-6_27

Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis. *IEEE Access*, *10*, 56720–56739. https://doi.org/10.1109/ACCESS.2022.3177752

Sharma, K. K., Tomar, M., & Tadimarri, A. (2023). Optimizing Sales Funnel Efficiency: Deep Learning Techniques for Lead Scoring. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (Online)*, *2*, 261–274. https://doi.org/10.60087/jklst.vol2.n2.p274

Sharma, M. (2022). *Identifying Factors Contributing to Lead Conversion Using Machine Learning to Gain Business Insights MSc Research Project MSCDADJAN22*.

Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2023). *A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU*. https://arxiv.org/abs/2305.17473v2

Song, Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*, 130–135. https://api.semanticscholar.org/CorpusID:18242585

Srinidhi, C. L., Ciga, O., & Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, *67*, 101813. https://doi.org/https://doi.org/10.1016/j.media.2020.101813

Suhaidi, M., Kadir, R. A., & Tiun, S. (2021). A REVIEW OF FEATURE EXTRACTION METHODS ON MACHINE LEARNING. *JOURNAL INFORMATION AND TECHNOLOGY MANAGEMENT (JISTM)*, *6*(22), 51–59. https://doi.org/10.35631/JISTM.622005

Sundman, S.-M. (2021). *Customer Acquisition Development Plan : Improving Lead Generation*. http://www.theseus.fi/handle/10024/356250

Swelsen, C. W. J. M. (2019). *Proposing a generic online lead scoring model for a B2C market*.

Thakur, A., & Konde, A. (2021). Fundamentals of Neural Networks. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, *9*, 2321–9653. www.ijraset.com

Thompson, N., Greenewald, K., Lee, K., & Manso, G. F. (2021). *The Computational Limits of Deep Learning*.

Vijayakumar, K., Kadam, V. J., & Sharma, S. K. (2021). Breast cancer diagnosis using multiple activation deep neural network. *Concurrent Engineering*, *29*(3), 275–284. https://doi.org/10.1177/1063293X211025105

Vo, N. N. Y., Liu, S., Li, X., & Xu, G. (2021). Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems*, *212*, 106586. https://doi.org/https://doi.org/10.1016/j.knosys.2020.106586

Wang Blippar, J., Xu Blippar, J., & Wang Blippar, X. (2018). *Combination of Hyperband and Bayesian Optimization for Hyperparameter Optimization in Deep Learning*. https://arxiv.org/abs/1801.01596v1

Wang, L., Ye, W., Zhu, Y., Yang, F., & Zhou, Y. (2023). Optimal parameters selection of back propagation algorithm in the feedforward neural network. *Engineering Analysis with Boundary Elements*, *151*, 575–596. https://doi.org/https://doi.org/10.1016/j.enganabound.2023.03.033

Waters, P. (2022). *Lead generation in Business-to Business marketing and sales : how can Heimo Films improve their lead generation process?* http://www.theseus.fi/handle/10024/753271

Wu, M., Andreev, P., & Benyoucef, M. (2024). The state of lead scoring models and their impact on sales performance. *Information Technology and Management*, *25*(1), 69–98. https://doi.org/10.1007/S10799-023-00388-W/TABLES/8

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Yang, G. R., & Wang, X.-J. (2020). Artificial Neural Networks for Neuroscientists: A Primer. *Neuron*, *107*(6), 1048–1070. https://doi.org/10.1016/j.neuron.2020.09.005

Yim, W. Y., Khaw, K. W., Lim, S. T., & Chew, X. (2024). Enhancing Conversions and Lead Scoring in Online Professional Education. *International Journal of Management, Finance and Accounting*, *5*(1), 15–63. https://doi.org/10.33093/IJOMFA.2024.5.1.2

Zabor, E. C., Reddy, C. A., Tendulkar, R. D., & Patil, S. (2022). Logistic Regression in Clinical Studies. *International Journal of Radiation Oncology*Biology*Physics*, *112*(2), 271–277. https://doi.org/10.1016/J.IJROBP.2021.08.007

Zelikovitz, S., & Hirsh, H. (2001). Using LSI for text classification in the presence of background text. *Proceedings of the Tenth International Conference on Information and Knowledge Management*, 113–118. https://doi.org/10.1145/502585.502605

Zhang, C., Cao, L., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. *Sustainable Cities and Society*, *39*, 508–518. https://doi.org/https://doi.org/10.1016/j.scs.2018.02.016

Zhang, G. (2023). Comparison of Machine Learning Models and Traditional Models for Forecasting in the Economy. *Computer Life*, *11*(3), 1–6. https://doi.org/10.54097/N11MBN72

Zhao, Y., Wang, T., Bove, R., Cree, B., Henry, R., Lokhande, H., Polgar-Turcsanyi, M., Anderson, M., Bakshi, R., Weiner, H. L., & Chitnis, T. (2020). Ensemble learning predicts multiple sclerosis disease course in the SUMMIT study. *Npj Digital Medicine 2020 3:1*, *3*(1), 1–8. https://doi.org/10.1038/s41746-020-00338-8

Zhu, F., Zhang, W., Chen, X., Gao, X., & Ye, N. (2023). Large margin distribution multi-class supervised novelty detection. *Expert Systems with Applications*, *224*, 119937. https://doi.org/10.1016/J.ESWA.2023.119937

Zhuang, H., Wang, X., Bendersky, M., & Najork, M. (2020). Feature Transformation for Neural Ranking Models. *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1649–1652. https://doi.org/10.1145/3397271.3401333

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)