International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 08 August-2026 | Impact Factor: 3.5

Design and Development of an Advanced Machine Learning Framework for Predicting B2b Sales Success

Mr. T. Narasimhulu[1]*, Dr. B. Poorna Satyannarayana[2] , Mr. S. Durga Prasad[3]
[1]PG Students, Department of CSE, BITS (A), Vizag, 530048
[2]Professor, Department of CSE, BITS (A), Vizag, 530048
[3]HoD, Department of CSE, BITS (A), Vizag, 530048

Abstract- The aims of this project are two-fold, firstly to help a Fortune 500 paper & paper product company encode what makes for sales success and secondly to create a model that will predict with a reasonable accuracy of a sales success. The ultimate goal is to boost sales close rates, reduce sales cycles and lower the cost of sales, thus generating the desired long-term benefit of the company generating additional revenue on top and additional profit at the bottom. The research team developed a number of models to predict the win probability for individual sales opportunities, selecting the model that most optimally predicted the win and was able to provide insights to serve as the core for a client tool. To do this, the team used the company's structured and unstructured data from its customer relationship management (CRM) system, Salesforce.com. They tried a variety of methods such as binomial logit, decision tree methods such as boosting with gradient boost and random forest. Individual attributes of customers, opportunities, and internal documentation methods that have the greatest influence on sales success were identified. The best model achieved a accuracy of 80%, precision of 86% and recall of 77% for predicting win propensity, which was an improvement over the current accuracy of sales predictions.

Keywords: B2B Sales; Machine Learning; Predictive Analytics; Sales Success Prediction; Customer Conversion; XGBoost; Random Forest.

How to Cite this Article:

Narasimhulu, T., Satyannarayana, B. P. & Prasad, S. D. (2026). Design and Development of an Advanced Machine Learning Framework for Predicting B2b Sales Success. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(8), 1-9. https://doi.org/10.55041/ijcope.v2i8.194

https://doi.org/10.55041/ijcope.v2i8.194

1. Introduction

Business-to-Business (B2B) sales is an essential component of organizational growth, revenue generation, customer acquisition, and market expansion. Unlike B2C transactions, B2B sales generally involve multiple decision-makers, lengthy sales cycles, technical evaluations, negotiations, customized requirements, and organizational approvals. Therefore, accurately predicting whether a sales opportunity will be successfully converted is a complex task. Traditional sales forecasting methods mainly depend on historical trends, salesperson experience, CRM records, and managerial judgment. While they are helpful, these techniques may fail to accurately reflect more complex relationships in large and evolving sales data. With the adoption of Customer Relationship Management (CRM) systems and digital sales platforms, a ton of customer and sales information has been created. The information in these data encompasses lead attributes, customer profiles, opportunity phases, quotation values, product data, sales engagements, salesperson actions, and transaction history. B2B sales analytics involves analyzing this data to find patterns, assess opportunities, predict results, and make business decisions. Machine learning is one way to work with a trove of past sales data and turn it into predictive insights and opportunities that are likely to succeed. There are a number of issues that you need to contend with in B2B sales prediction. Purchase cycles can be long and go through many phases, and may have multiple decision makers. Sales data may also be heterogeneous, including numerical, categorical, text and temporal features. Absence of values, inconsistent records, duplicate records, and noisy data may have a detrimental impact on model performance. Class imbalance is another issue that needs to be considered as well since there can be a large number of unsuccessful opportunities that exceed the number of successful opportunities. So, accuracy by itself might not be a good indicator of the models' performance. Indicators like precision, recall, F1-score and ROC-AUC are thus needed. Further, shifts in customer behavior, market conditions, competition and business strategies can diminish the ability of models formed from past data to generalize to future data. Machine learning can provide insights into patterns and relationships that would not otherwise be apparent in sales data, connecting sales attributes with past performance. Supervised learning algorithms could be trained with the past sales opportunities that were successful or unsuccessful and then used on new opportunities. Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, Decision Tree, Random Forest and XGBoost are the algorithms that can be utilized for predicting sales success in B2B. Logistic

Regression gives an easily-interpreted baseline model, and SVM can model more complicated decision boundaries. Decision Tree and Random Forest algorithms are good for classification and feature analysis, while XGBoost can reveal nonlinear relationships and interactions among features. Multiple algorithms can be compared to determine the best model for a specific B2B sales dataset. The traditional B2B sales forecasting methods mostly rely on salesperson's best guess, historical averages, and manually managed sales pipelines. These methods may not be sufficient to fully represent the interactions between different types of customer characteristics, sales activities, opportunity attributes and purchasing behavior. While machine-learning techniques have proven to be very promising, an integrated system that encompasses data preprocessing, feature engineering, predictive modeling, model optimization, and model comparisons is still needed for B2B sales-success prediction. This research's main purpose is to give data-centric B2B sales decision making. Organisations spend lots of time and money on lead generation and opportunity management and not all opportunities have the same chance of conversion. A reliable predictive model can aid sales teams to determine which opportunities they should target, prioritize leads, allocate resources appropriately, and enhance sales planning. With its growing volume, CRM information continues to drive the need for machine-learning tools that can transform past information into predictive insights and intelligence. The key goal is to create an advanced machine learning system to predict B2B sales success. The particular aim is to: To build a sophisticated machine learning model to forecast the success of a B2B sale, taking out the noise from the sales data and extracting relevant features to drive sales. To build and compare the performance of predictive models (SVM, Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, XGBoost) using appropriate performance measures to determine which is the best predictive model. To implement the developed predictive framework to B2B lead prioritization and data-driven sales decision making, which is going to enhance sales opportunity management and forecasting. The research adds an integrated machine-learning framework, which incorporates data preprocessing, feature engineering, model development, evaluation, and prediction of B2B sales success. It compares different machine learning models and determines key factors that impact sales results. The framework can be used to help sales teams prioritize opportunities, allocate resources, enhance forecasting, and make informed decisions. The intent of the research is to bring together high-end machine-learning tools and real-world B2B sales management use.

2. Literature Review

As a result of the ability to detect complex patterns in historical sales data, machine learning (ML) has increasingly been used in sales predictions. In contrast to the conventional forecasting models that primarily rely on historical trends and statistical assumptions, the models using ML allow incorporating customer, product, transaction, and the sales-process attributes. Yan et al. proved that machine learning can be used to predict sales in the sales pipeline; and later researchers have investigated logistic regression, support vector machine (SVM), decision tree, random forest and gradient boosting methods for forecasting sales. The added value of predictive analytics in B2B sales lies in the fact that B2B deals are intricate, involve multiple parties, and have longer sales cycles. Rezazadeh showed the viability of generalized machine-learning methods for the B2B sales predictive modeling. Likewise, structured sales data have been applied to supervised learning to predict B2B sales. These methods help with opportunity management, sales forecasting, opportunity targeting, and resource allocation. Lead conversion prediction relies on customer attributes, engagement, lead source, opportunity stage, and interaction histories to predict conversion likelihood. The latest B2B lead-scoring studies demonstrate the importance of ML to prioritize good leads. Customer classification is another tool that aids segmentation and customized marketing approaches. While there are some successful studies reported, most of those studies target specific algorithms, datasets, industries, and individual forecasting tasks. There are scarce studies that compare the performance of several ML algorithms for predicting sales success in B2B environments. Thus, a comprehensive approach that integrates the preprocessing, feature engineering, comparative modeling, and performance evaluation is needed.

Table 1: Comparative Analysis of Existing Research

Ref.	Year	Study / Authors	Research Focus	Method / Algorithm	Key Contribution	Limitation / Research Gap
[1]	2015	Yan et al.	Predictive sales-pipeline analytics	Hawkes Process / ML	Predicted sales lead win-propensity using seller–lead interactions and temporal behavior.	Focused on a specific enterprise sales environment; broader ML comparison was limited.
[2]	2018	Ali & Lee	CRM sales prediction	Continuous time-evolving classification	Used CRM data to identify salesmanship patterns and success factors.	Primarily focused on CRM and time-evolving classification.
[3]	2019	Mortensen et al.	B2B sales-success prediction	Machine Learning	Directly investigated the prediction and definition of B2B sales success.	Scope and dataset were specific to the studied organization.
[4]	2019	Harrison & Ajjan	CRM technology and marketing practice	CRM / Analytics	Examined the relationship between	More focused on CRM adoption and practice than predictive modeling.

					CRM technology and marketing practice.	
[5]	2019	Yan Qi et al.	E-commerce sales forecasting	Deep Neural Network	Developed a deep-learning framework for sales forecasting.	E-commerce oriented rather than B2B sales-success prediction.
[6]	2020	Rezazadeh	B2B sales predictive modeling	Azure ML / Ensemble ML	Proposed a generalized workflow for B2B predictive modeling.	Further investigation is required across different B2B datasets and algorithms.
[7]	2020	Calixto & Ferreira	B2B salesperson performance	Predictive Analytics	Applied predictive analytics to evaluate salesperson performance.	Focused mainly on salesperson performance rather than overall sales-success prediction.
[8]	2022	Rohaan et al.	B2B sales forecasting	Supervised ML / NLP	Applied supervised ML to spare-parts sales forecasting in an after-sales B2B environment.	Case-specific and focused on a particular product/service context.
[9]	2022	Bi et al.	Sales forecasting accuracy	Tensor Factorization	Improved forecasting by incorporating demand-awareness information.	Primarily concerned with forecasting rather than sales-opportunity success classification.
[10]	2022	Fries & Ludwig	ML-based sales forecasting	Machine Learning	Investigated socio-technical challenges in implementing ML sales forecasting.	Focused strongly on organizational and implementation challenges.
[11]	2023	Martins & Galegale	Sales forecasting	ML Algorithms	Examined machine-learning algorithms for sales forecasting.	General sales forecasting; limited focus on B2B opportunity-level prediction.
[12]	2024	Shi et al.	CRM and machine learning	ML / CRM Analytics	Investigated ML applications for improving CRM-related activities.	Requires further development toward integrated B2B sales-success prediction.
[13]	2025	Abaddi	Sales management and CRM	CRM / Simulation	Examined improvement of sales management through CRM and simulation.	Not primarily an ML-based B2B sales-success framework.
[14]	2025	González-Flores et al.	B2B lead prioritization	15 Classification Algorithms	Developed an ML-based B2B lead-scoring model; Gradient Boosting performed best among the tested classifiers.	Case study based on a specific B2B software company; broader framework validation remains necessary.
[15]	2025	Henna et al.	B2B CRM and sales forecasting	Spectral Graph Convolutional Networks	Applied graph-based deep learning to model CRM relationships and sales forecasting.	Advanced architecture may require substantial data and computational resources.
[16]	2025	Habel et al.	Sales-pipeline technology	Automated Lead Nurturing	Investigated automated technology for improving sales-pipeline lead nurturing.	Focuses on lead nurturing rather than comprehensive sales-success prediction.
[17]	2025	Sustainable Supply Chain Study	Sales prediction	Machine Learning	Applied ML to sales prediction within sustainable supply-chain forecasting.	Broader supply-chain context rather than specifically B2B sales success.
[18]	2025	B2B Automobile Sales Study	B2B automobile sales forecasting	ANN-GA	Applied ANN and Genetic Algorithm techniques to B2B automobile sales forecasting.	Industry-specific and focused primarily on forecasting.

The research shows that there's been great strides in using machine learning for sales forecasting, CRM analytics, lead scoring, and customer classification. There are, however, a number of gaps in research. Existing work often focuses on specific aspects of B2B sales, such as data preprocessing, feature engineering, multiple classification algorithms, model evaluation, and prediction of sales success, without combining them into one framework. Secondly, datasets and evaluation methods are not necessarily comparable and direct comparison of algorithms is not easy. Third, class imbalance, feature relevance, shift in customer behaviour, and model interpretability are not always considered.

3. Proposed Methodology

The methodology proposed in this paper is a machine learning-based framework that integrates the analysis of historical customer, lead, and opportunity information along with sales data to predict the success of a B2B sale. The framework comprises of successive phases such as data collection, data preprocessing, feature engineering, feature selection, model development, model training, model validation, hyperparameter optimization, and prediction. The main goal is to convert raw B2B sales information into actionable predictive data that can aid in lead prioritization and data-driven sales decisions. The model is a binary classification problem where each sales opportunity is labeled as a success or failure. Several machine-learning algorithms are created and tested to find the best machine-learning model for the prediction task. The proposed system architecture includes five major layers of the system, namely the data acquisition layer, data preparation layer, predictive modelling layer, evaluation layer, and decision support layer. The data acquisition layer gathers data from the past from the CRM and sales systems. Data preparation layer is responsible for cleaning, transformation, encoding, feature engineering and feature selection. The predictive modeling layer creates multiple machine-learning models. The evaluation layer provides the evaluation of the models based on suitable metrics. Lastly, the decision support layer generates predictions about the likelihood of success of new B2B sales opportunities based on the chosen model. Sales data that are relevant to B2B are gathered from available CRM databases, organizational sales records, records of customer interactions and other appropriate datasets. Lead source, Customer type, Industry, Geographical region, Sales stage, deal value, number of interactions, salesperson activity, product category, previous purchase behavior, and sales-cycle duration are some of the important attributes. Target variable is the ultimate sales result. As part of the data set preparation, data quality, relevance, completeness, and consistency are taken into account. Raw sales data typically have missing data, duplicated data, inconsistent data formats, categorical data, and possible outliers. Data pre-processing is hence done prior to the model development. Duplicate and irrelevant records are deleted, missing values are dealt with appropriately, and categorical variables are encoded into a numerical representation using an appropriate technique. Where necessary, numerical features might be normalized or standardized. The data set is then split into training and testing sets with proper class distribution. Feature engineering is the process of creating features that are useful for prediction. These features could encompass frequency of customer interactions, time of last sales opportunity, intensity of interaction, length of sales cycle, average transaction value, number of purchases in the past, conversion rate history, etc. Additional information from individual raw variables may be available from derived features. Feature engineering can enhance the capacity of a model to learn and the prediction results. Feature Selection identifies the most relevant features for predicting sales success. Unimportant and unnecessary attributes can add noise and complexity to the computations. Various techniques for feature-importance and model-based feature-selection methods as well as statistical analysis and correlation analysis can be used to find predictors that are significant. Choosing informative features also results in better interpretability of the model. The following six machine-learning algorithms are taken into account: Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, Decision Tree, Random Forest and XGBoost. These models are various learning strategies and serve as a general foundation for comparison. Logistic Regression provides a baseline interpretable model, and tree-based and boosting algorithms can learn nonlinear relationships between B2B sales features. Selected data set is split to train and test sets. The training set is employed to build up the predictive models and the testing set is set aside for independent evaluation of the predictive models. Cross validation can be used to check the model stability and minimize overfitting. Metrics of model performance include accuracy, precision, recall, F1-score, confusion matrix and ROC-AUC. Hyperparameter optimization is used to get better performance of the model. The algorithm determines the optimization of parameters like tree depth, number of estimators, learning rate, kernel parameters, strength of regularization, and minimum sample requirements. The optimization methods of Grid Search, Random Search etc. can be used. Optimised parameters are selected according to the validation performance without overfitting. The best performing model is used after training and optimization based on results from an extensive evaluation. The chosen model is fed with details of a new B2B sales opportunity and forecasts the probability of success. The prediction can be conveyed as a classification label and, if available, as a probability value. High-probability opportunities can be prioritized for immediate sales attention, while lower-probability opportunities can be managed through appropriate follow-up strategies. The proposed methodology therefore establishes a systematic connection between B2B sales data and actionable predictive intelligence, enabling organizations to improve opportunity prioritization, sales forecasting, resource allocation, and data-driven decision-making.

International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 08 August-2026 | Impact Factor: 3.5

4. Dataset Collection

Table 2: B2B Sales Dataset Attributes

No.	Attribute / Feature	Description	Data Type	Values / Categories
1	Product	Product associated with the sales opportunity	Categorical	Product A, Product B, Product C, Product D, Product F, ...
2	Seller	Salesperson responsible for the opportunity	Categorical	Seller 1, Seller 2, Seller 3, ..., Seller 20

No.	Attribute / Feature	Description	Data Type	Values / Categories
1	Product	Product associated with the sales opportunity	Categorical	Product A, Product B, Product C, Product D, Product F, ...
2	Seller	Salesperson responsible for the opportunity	Categorical	Seller 1, Seller 2, Seller 3, ..., Seller 20
3	Authority	Level of decision-making authority of the customer	Categorical	Low, Mid, High
4	Company Size	Size/category of the customer organization	Categorical	Small, Mid, Big
5	Competitor	Presence of competitor involvement	Categorical	Yes, No
6	Purchase Dependency	Customer dependency on the purchase	Categorical	Yes, No, Unknown
7	Partnership	Existing partnership with the customer	Categorical	Yes, No
8	Budget Allocation	Availability of allocated budget	Categorical	Yes, No, Unknown
9	Formal Tender	Whether a formal tender process is involved	Categorical	Yes, No
10	RFI	Request for Information involvement	Categorical	Yes, No
11	RFP	Request for Proposal involvement	Categorical	Yes, No
12	Growth	Customer/company growth status	Categorical	Growth, Stable
13	Position Status	Customer's current position/status	Categorical	Neutral, Yes
14	Source	Source through which the opportunity was generated	Categorical	Referral, Joint Partner, Online, Event, ...
15	Client	Existing or new client classification	Categorical	Current, New
16	Scope	Clarity of project/opportunity scope	Categorical	Clear, Few Questions, Unclear
17	Strategic Deal	Strategic importance of the deal	Categorical	Very Important, Important, Average, Unimportant
18	Cross-Sale	Cross-selling opportunity	Categorical	Yes, No
19	Up-Sale	Up-selling opportunity	Categorical	Yes, No
20	Deal Type	Type/category of sales deal	Categorical	Project, Maintenance, Solution, Service
21	Needs Detail	Availability/clarity of customer requirements	Categorical	Yes, No, Information Gathering
22	Attention / Client Status	Customer attention or engagement level	Categorical	Strategic, Normal, First Deal, Info Gathering
23	Status (Target)	Final sales outcome (target variable)	Categorical	Won, Lost

No | Yes | Yes | Growth | Neutral | Current

Product	Seller	Authority	Comp. Size	Competitor	Purch. Dep.	Partnership	Budget Alloc.	Formal Tender	RFI	RFP	Growth	Posit. Status	Source	Client	Scope	Strategic Deal	Cross-Sale	Up-Sale	Deal Type	Needs Detail	Client Status	Status (Target)
Product A	Seller 6	Mid	Mid	No	Unknown	No	Yes	No	Yes	Yes	Growth	Neutral	Current	Current	Clear	Very Important	No	Yes	Project	Yes	Strategic	Won
Product B	Seller 8	Mid	Big	No	Yes	No	Unknown	No	No	Yes	Stable	Yes	Referral	Current	Clear	Unimportant	No	Yes	Maintenance	Yes	Strategic	Won
Product F	Seller 1	Mid	Big	No	No	Yes	No	No	No	Yes	Stable	Yes	Joint Partner	Current	Few Questions	Average	No	No	Project	Info Gathering	Strategic	Won
Product B	Seller 8	Mid	Big	No	Yes	No	No	No	No	No	Growth	Yes	Joint Partner	Current	Clear	Very Important	No	Yes	Maintenance	Yes	Strategic	Won
Product C	Seller 1	Mid	Small	Unknown	No	No	No	No	No	No	Stable	Neutral	Unknown	New	Few Questions	Average	No	No	Project	No	Normal	Won
Product B	Seller 20	Mid	Big	Unknown	Unknown	No	No	No	No	No	Growth	Yes	Referral	New	Clear	Very Important	No	No	Project	Yes	First Deal	Lost
Product B	Seller 1	Mid	Mid	Yes	Unknown	Yes	Unknown	No	No	Yes	Stable	Neutral	Referral	New	Clear	Very Important	No	No	Project	Yes	First Deal	Won

Table 3: Hepatitis Dataset Attributes

S. No.	Attribute / Feature	Description	Data Type	Values / Categories
1	**Class**	Indicates the patient's disease/outcome class	Categorical (Target)	1, 2
2	**Age**	Age of the patient in years	Numerical	Range: 7 – 78
3	**Sex**	Gender of the patient	Categorical	1 = Male, 2 = Female
4	**Steroid**	Indicates whether steroid treatment was administered	Categorical	1 = Yes, 2 = No
5	**Antivirals**	Indicates whether antiviral treatment was administered	Categorical	1 = Yes, 2 = No
6	**Fatigue**	Presence or absence of fatigue	Categorical	1 = Yes, 2 = No
7	**Malaise**	Presence or absence of malaise	Categorical	1 = Yes, 2 = No
8	**Anorexia**	Presence or absence of anorexia	Categorical	1 = Yes, 2 = No
9	**Liver Big**	Indicates whether the liver is enlarged	Categorical	1 = Yes, 2 = No
10	**Liver Firm**	Indicates whether the liver is firm	Categorical	1 = Yes, 2 = No
11	**Spleen Palpable**	Indicates whether the spleen is palpable	Categorical	1 = Yes, 2 = No
12	**Ascites**	Presence or absence of ascites	Categorical	1 = Yes, 2 = No
13	**Varices**	Presence or absence of varices	Categorical	1 = Yes, 2 = No
14	**Bilirubin**	Level of bilirubin in the blood	Numerical	Range: 0.3 – 2.8
15	**Alk Phosphate**	Alkaline phosphatase level	Numerical	Range: 26 – 295
16	**SGOT**	Serum glutamic-oxaloacetic transaminase level	Numerical	Range: 14 – 648
17	**Albumin**	Albumin concentration in the blood	Numerical	Range: 2.1 – 4.9
18	**Protime**	Prothrombin time	Numerical	Range: 10 – 100
19	**Histology**	Histological examination result	Categorical	1 = Normal, 2 = Abnormal

Class	Age	Sex	Steroid	Antivirals	Fatigue	Malaise	Anorexia	Liver Big	Liver Firm	Spleen Palpable	Ascites	Varices	Bilirubin	Alk Phosphate	SGOT	Albumin	Protime	Histology
2	30	2	1	1	2	2	2	2	2	2	2	2	0.9	85	18	4.5	61	1
2	50	1	1	2	1	2	2	2	2	2	2	2	0.9	135	42	3.5	61	1
2	78	1	1	2	2	2	2	2	2	2	2	2	0.7	96	32	4.0	61	1
2	34	1	2	2	1	2	2	2	2	2	2	2	0.9	95	28	4.0	75	1
2	51	1	1	2	2	2	2	1	2	2	2	2	1.42	105	85	3.81	61	1
6	23	1	2	2	1	2	2	2	2	2	2	2	1.7	105	85	3.81	61	1
2	39	1	2	2	1	2	2	2	2	2	2	2	0.9	105	48	4.0	61	1
2	38	1	2	2	1	2	2	2	2	2	2	2	1.3	78	30	4.4	85	1
2	27	2	1	1	2	2	2	2	1	2	2	2	1.3	59	249	3.7	54	1
2	61	1	1	2	1	1	2	2	1	2	2	2	1.1	81	60	3.9	52	1

Data is stored in one of the various data structures. In a database, for example, a data set might contain a collection of business data (names, salaries, contact information, sales figures, and so forth). The database itself can be considered a data set, as can bodies of data within it related to a particular type of information, such as sales data for a particular corporate department. The term data set was coined by IBM and had a similar meaning as file. In an IBM mainframe operating system, a data set s a named collection of data that contains individual data units organized (formatted) in a specific, IBM-prescribed way and accessed by a specific access method based on the data set organization. There are three types of data set organization: sequential, relative sequential, indexed sequential and partitioned. Various access methods are used: Virtual Sequential Access Method (VSAM); Indexed Sequential Access Method (ISAM). Data sets have several features that determine their structure and properties. These include attribute/variable numbers and types, and statistics that apply to them, including standard deviation and kurtosis. The values can be numeric (e.g. real numbers, integers, etc.) and may be a person's height in cm, or they can be nominal (i.e. not numeric values) and may be a person's ethnicity. Values can be of any of the types listed as a level of measurement, more generally. The values for each variable are typically all the same type. There can also be missing values, however, and these should be represented in some fashion. Typically, data sets are obtained from real observations made of a sample of a statistical population, with each row representing the observations made of one element of the statistical population. Other algorithms can also produce data sets, which are then used to test some types of software. Some modern statistical analysis software such as SPSS still presents their data in the classical data set fashion. If data is missing or suspicious an imputation method may be used to complete a data set.

5. Machine Learning Models

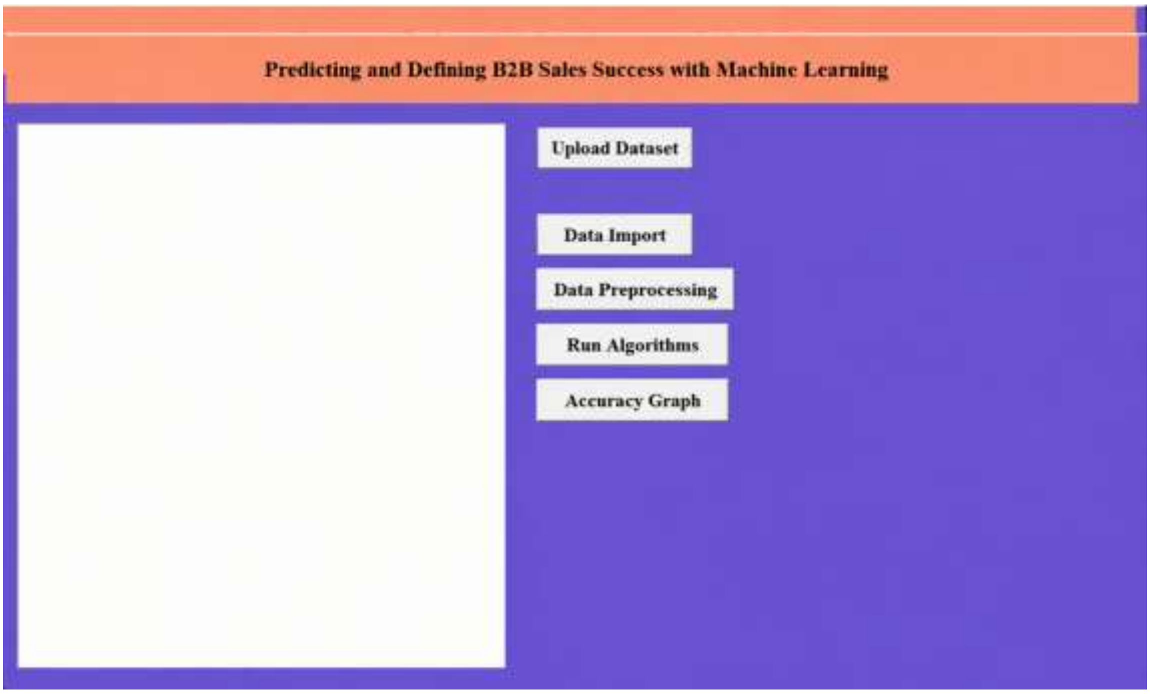
By using a machine learning algorithm, you will gain a practical and in-depth understanding of the algorithm. This knowledge can also be used to internalize the mathematical description of the algorithm: thinking of the vectors and matrices as arrays, and thinking of the intuition for the computation of the transformations on those structures. When implementing a machine learning algorithm, there are many micro-decisions to be made and these are usually not included in the formal algorithm description. Learning and parameterizing these decisions can quickly catapult you to intermediate and advanced level of understanding of a given method, as relatively few people make the time to implement some of the more complex algorithms as a learning exercise. Mastery: First step towards mastering an algorithm is implementation of it. When implementing an algorithm, you must be very familiar with it. You are also creating your own laboratory for tinkering to help you internalize the computation it performs over time, such as by debugging and adding measures for assessing the running process. Production Systems: Custom implementations of algorithms are typically required for production systems because of the changes that need to be made to the algorithm for efficiency and efficacy reasons. Ultimately, better, faster and less resource intensive can translate to lower business cost and higher revenue, and working with hand-coded algorithms will help you learn how to provide these solutions. Literature Review: When implementing an algorithm you are performing research. One has to find and read several formal and canonical descriptions of the algorithm. You will also find other implementations of the algorithm and do a code review to ensure your understandings. You are performing targeted research, and learning how to read and make practical use of research publications.

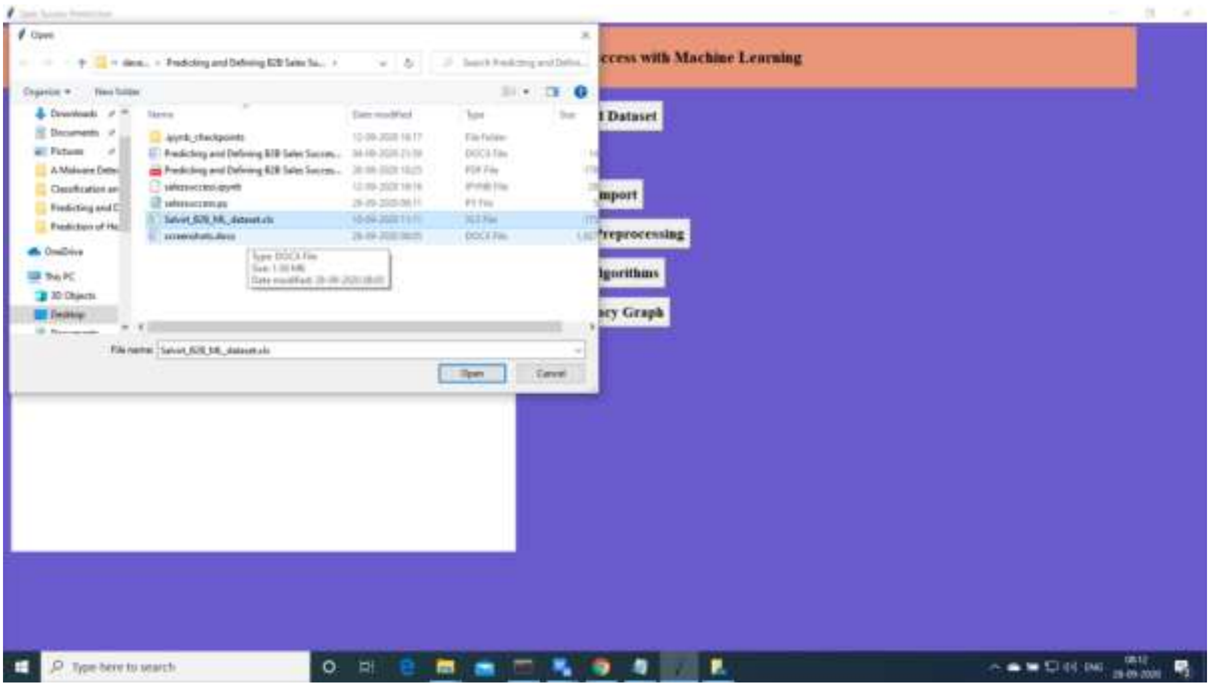
6. Steps for Implementation

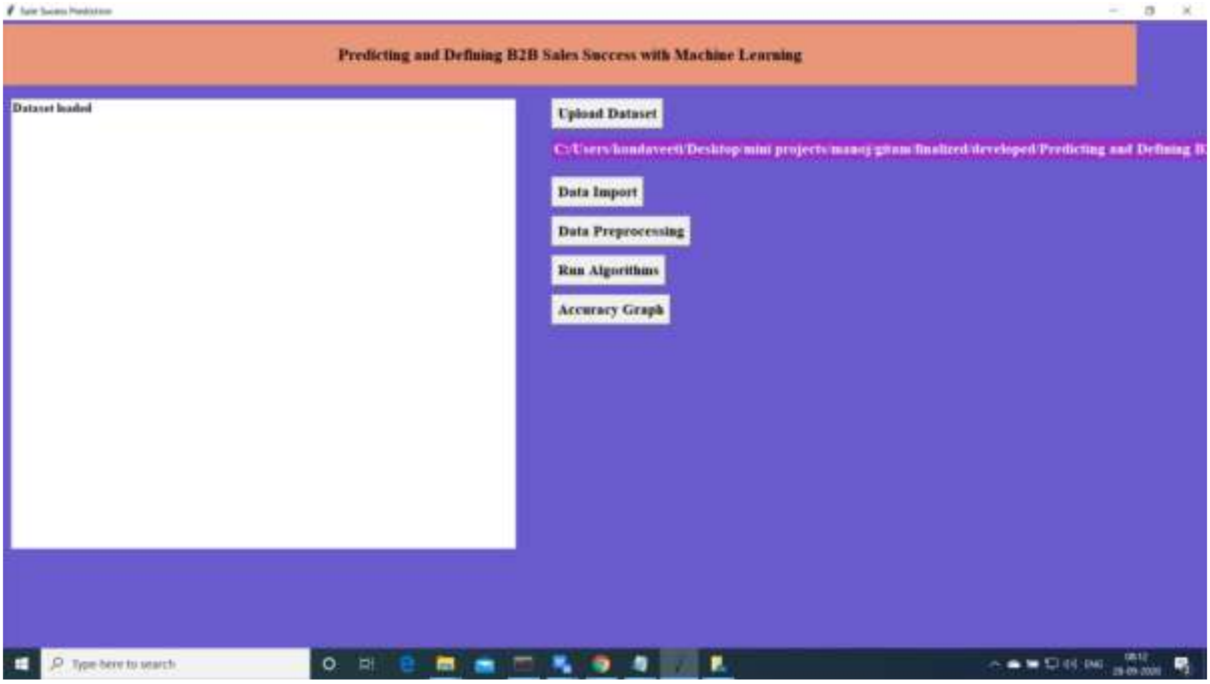
Step 1: Install the latest version of Anaconda (previously given on the information page).
Step 2: Open anaconda Prompt
Step 3: Conda create -n tf python=3.7
Step 4: Conda env list to activate the Conda environment, you can use the command:
Step 5: Install require software
scikit-image==0.17.2
scikit-learn==0.23.2
pandas==1.1.1
matplotlib==3.3.1
Pillow==7.2.0 29
plotly==4.10.0
opencv-python==4.4.0.42
spacy==2.3.2
lightgbm==3.0.0
mahotas==1.4.11
matplotlib==3.3.1
lightgbm==3.0.0
mahotas==1.4.11
nltk==3.5
matplotlib==3.3.1
xgboost==1.2.0
Jupyter
Step6: Activate environment for jupyter notebook (For execute the in jupter notebook) python -m ipykernel install --user --name=
Step 7: Goto project Directory
Project Development Modules:
Data Collection: Gather adequate samples of data and software.
Data Preprocessing: Do useful data processing on the sample, and take out the features.

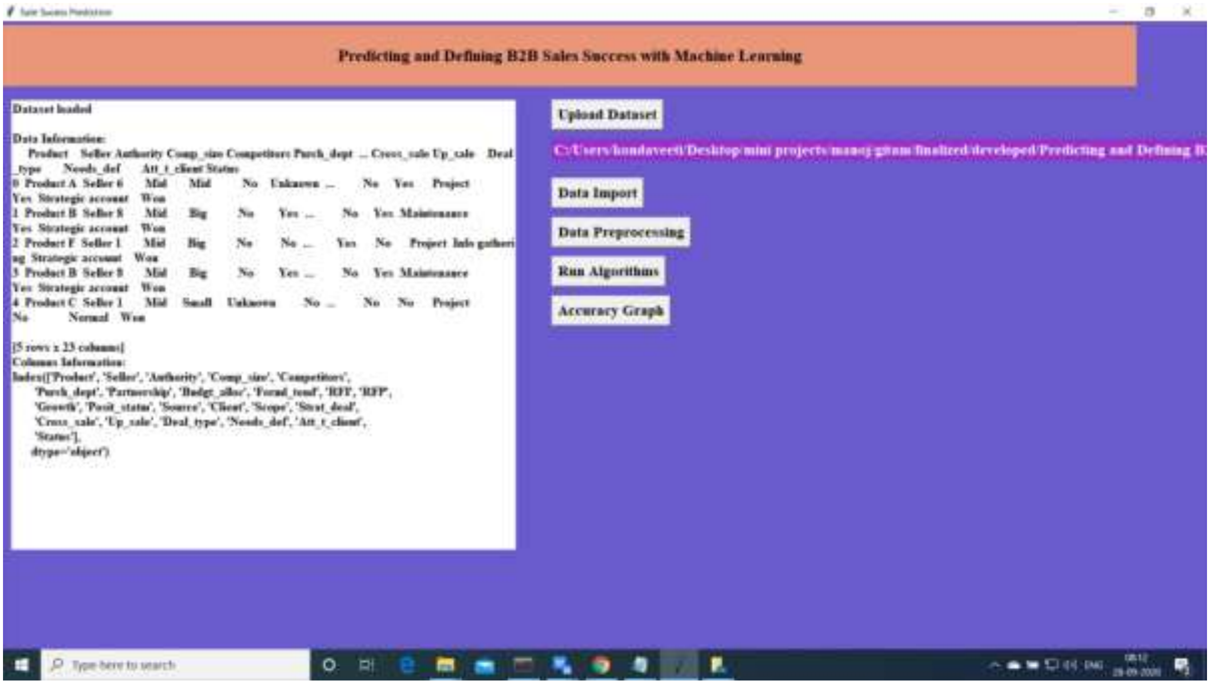
Train and Test Modelling: Split the data into train and test data Train will be used for training the model and Test data to check the performance.
Feature Selection: Select more features to classify.
Modelling: SVM Navie bayes, Random Forest, KNN, GBoost, Gradient boosting Decision tree, Ada boost with random forest. Use the machine learning algorithms to fuse the training and build the classification model.

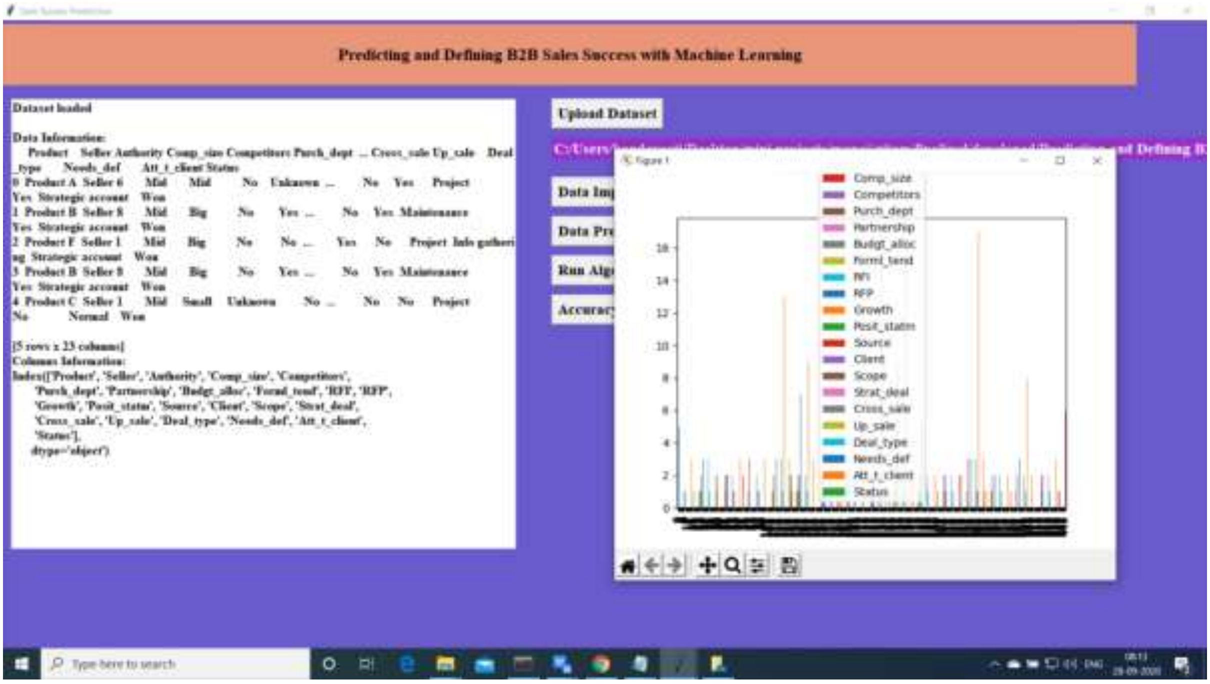
7. Results and Discussions

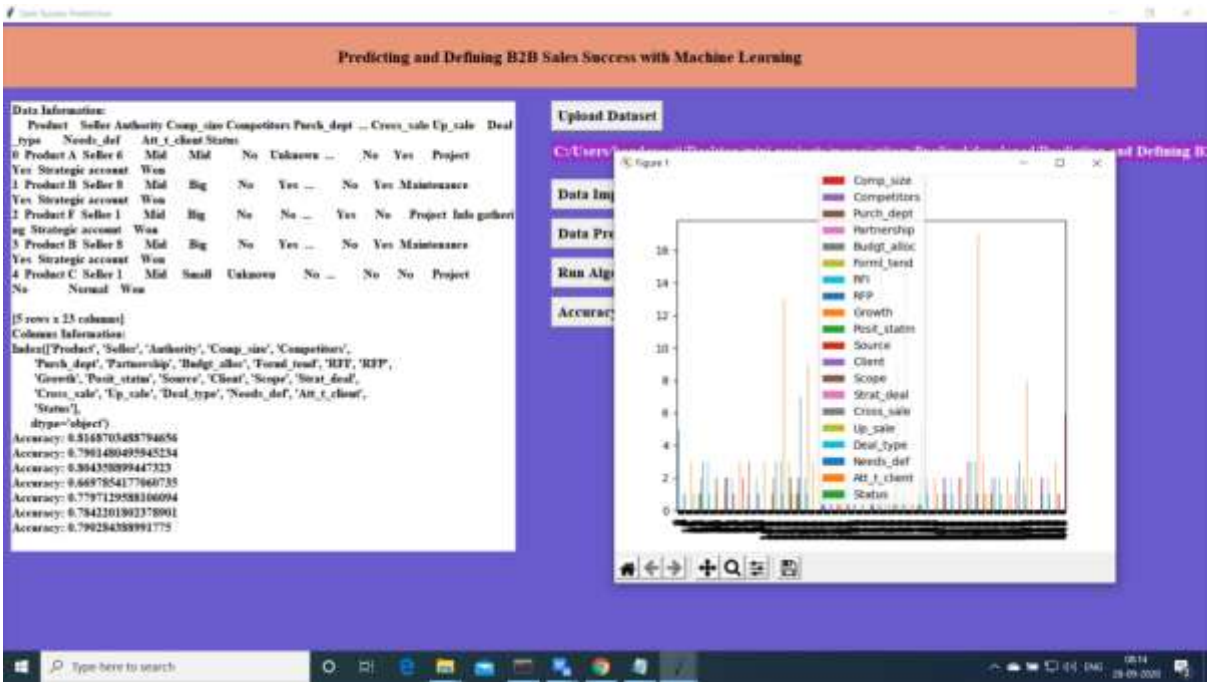




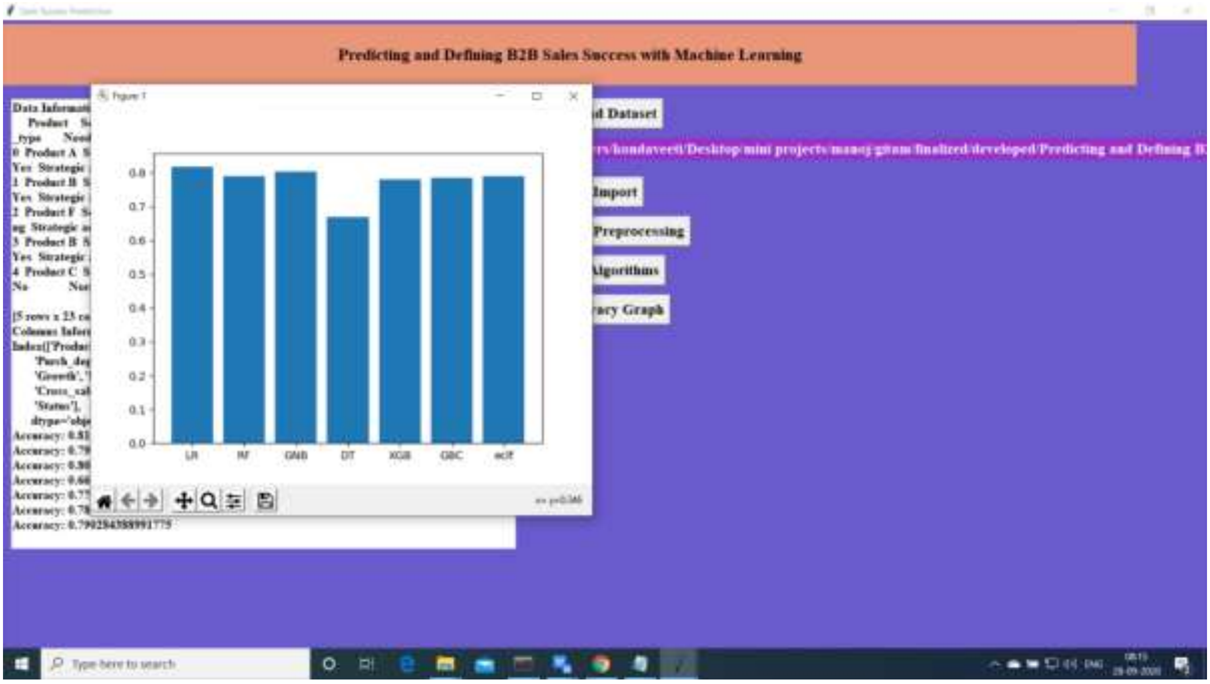

Predicting and Defining B2B Sales Success with Machine Learning
Dataset loaded
Upload Dataset
Data Import
Data Preprocessing
Run Algorithms
Accuracy Graph


Predicting and Defining B2B Sales Success with Machine Learning
Dataset loaded
Data Information:
[5 rows x 23 columns]
Columns Information:
Upload Dataset
Data Import
Data Preprocessing
Run Algorithms
Accuracy Graph





Accuracy Comparison for all the models



Extension is Logistic Regression algorithm performed well

Conclusion

This paper presented an initial step toward operationalizing predictive modeling for sales success within a large Fortune 500 paper and packaging company. Despite data inconsistencies, the results demonstrated promising predictive accuracy and revealed important factors influencing sales success. Future improvements should focus on enhancing data quality and volume, developing additional features, and evaluating advanced techniques such as neural networks. The study also highlights the importance of capturing opportunity changes over time through periodic snapshots. Accurate predictions could support sales-resource allocation, opportunity prioritization, and improved aggregate forecasting. Successful implementation will require continuous model refinement, data-management processes, regular updates, and organizational collaboration.

References

[1]. Yan, J., et al. "On Machine Learning towards Predictive Sales Pipeline Analytics." *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 1945–1951. https://doi.org/10.1609/aaai.v29i1.9455

[2]. Ali, M., and Y. Lee. "CRM Sales Prediction Using Continuous Time-Evolving Classification." *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. https://doi.org/10.1609/aaai.v32i1.11418

[3]. Mortensen, S. V., et al. "Predicting and Defining B2B Sales Success with Machine Learning." *2019 IEEE Symposium on Industrial Electronics & Applications (SIEDS)*, IEEE, 2019. https://doi.org/10.1109/SIEDS.2019.8735638

[4]. Harrison, D. E., and H. Ajjan. "CRM Technology: Bridging the Gap between Marketing Education and Practice." *Journal of Marketing Analytics*, vol. 7, 2019, pp. 112–123. https://doi.org/10.1057/s41270-019-00063-6

[5]. Yan Qi, Chenliang Li, Han Deng, Min Cai, Yunwei Qi, Yuming Deng. 2019.A Deep Neural Framework for Sales Forecasting in E-Commerce. In Pro-ceedings of 28th ACM International Conference on Information and Knowl-edge Management (CIKM'19). ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3357384.3357883

[6]. Rezazadeh, A. "A Generalized Flow for B2B Sales Predictive Modeling: An Azure Machine-Learning Approach." *Forecasting*, vol. 2, no. 3, 2020, pp. 267–283. https://doi.org/10.3390/forecast2030015

[7]. Calixto, N.; Ferreira, J. Salespeople Performance Evaluation with Predictive Analytics in B2B. *Appl. Sci.* **2020**, *10*, 4036. https://doi.org/10.3390/app10114036

[8]. Rohaan, D., et al. "Using Supervised Machine Learning for B2B Sales Forecasting: A Case Study of Spare Parts Sales Forecasting at an After-Sales Service Provider." *Expert Systems with Applications*, vol. 188, 2022, article 115925. https://doi.org/10.1016/j.eswa.2021.115925

[9]. Bi, X., et al. "Improving Sales Forecasting Accuracy: A Tensor Factorization Approach with Demand Awareness." *INFORMS Journal on Computing*, 2022. https://doi.org/10.1287/ijoc.2021.1147

[10]. Fries, M., and T. Ludwig. "Why Are the Sales Forecasts So Low? Socio-Technical Challenges of Using Machine Learning for Forecasting Sales in a Bakery." *Computer Supported Cooperative Work*, 2022. https://doi.org/10.1007/s10606-022-09458-z

[11]. Martins, E., and N. V. Galegale. "Sales Forecasting Using Machine Learning Algorithms." *Revista de Gestão e Secretariado*, vol. 14, no. 7, 2023. https://doi.org/10.7769/gesec.v14i7.1670

[12]. Shi, X., et al. "Revolutionizing Market Surveillance: Customer Relationship Management with Machine Learning." *PeerJ Computer Science*, 2024. https://doi.org/10.7717/peerj-cs.2583

[13]. Abaddi, S. (2025). Analysis and improvement of sales management through CRM and simulation utilisation. *Applied Operations and Analytics*, *1*(1). https://doi.org/10.1080/29966892.2024.2434267

[14]. González-Flores, L., et al. "The Relevance of Lead Prioritization: A B2B Lead Scoring Model Based on Machine Learning." *Frontiers in Artificial Intelligence*, 2025. https://doi.org/10.3389/frai.2025.1554325

[15]. Henna, S., et al. "Optimizing B2B Customer Relationship Management and Sales Forecasting with Spectral Graph Convolutional Networks: A Quantitative Approach." *Quantitative Finance and Economics*, 2025. https://www.aimspress.com/article/doi/10.3934/QFE.2025015

[16]. Habel, J., Hartmann, N. N., Wiseman, P., Ahearne, M. J., & Vaid, S. (2025). Sales Pipeline Technology: Automated Lead Nurturing. *Journal of Marketing*, *89*(5), 66-87. https://doi.org/10.1177/00222429251321417

[17]. "Enhancing Sustainable Supply Chain Forecasting Using Machine Learning for Sales Prediction." *Procedia Computer Science*, 2025, Volume 252, 2025, Pages 470-479. https://doi.org/10.1016/j.procs.2025.01.006

[18]. "B2B Automobile Sales Forecasting with Participatory Predictor Using ANN-GA." *Journal of Business & Industrial Marketing*, 2025, Volume 40, Issue 11, 31 October 2025, Pages 2077-2098. https://doi.org/10.1108/JBIM-04-2024-0274