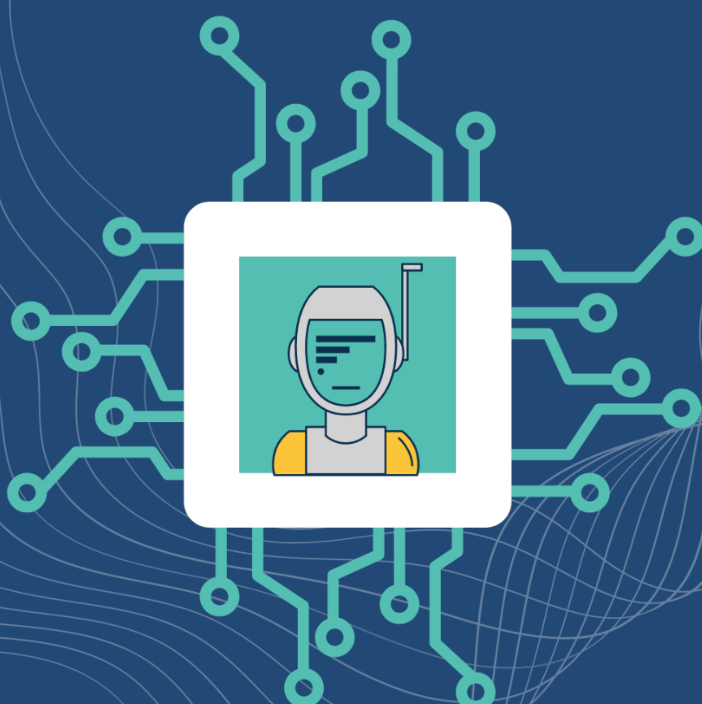
# AI-ENABLED SALES PLANNING

## OPPORTUNITY CLASSIFICATION



**Sales Opportunity Classification** **2**
**Abstract** **2**
**Introduction** **2**
Motivation 2
Goal 2
Overall Approach 2
**Background** **3**
**Detailed Approach** **3**
Data Collection 3
Exploratory Data Analysis 4
Statistical Analysis 4
Visualization 4
Data Preparation 4
Modeling 5
Model Evaluation 6
Hyper Parameter Tuning 6
Deployment 6
**Results Explanation** **7**
**Use Case** **7**
**Conclusions** **8**
Summary 8
Limitations 8
Future Work 8
**References** **8**

# Sales Opportunity Classification

# Abstract

This project aims to help salespeople sell more efficiently and effectively by leveraging artificial intelligence and machine learning technologies to streamline the prospecting process. By automating tasks relating to analyzing opportunity data and making inferences and identifying patterns to help classify opportunities as potential Win / Loss, salespeople can spend less time on prospecting and more time building relationships with customers and closing deals. This can reduce overhead costs, alleviate stress and pressure on sales teams, and ultimately drive business growth and success.

# Introduction

## Motivation

The sales industry is highly competitive, and companies are constantly looking for ways to improve the efficiency and effectiveness of their sales teams. One of the main challenges that salespeople face is prospecting, which can be time-consuming, expensive, and often result in low-quality leads.

1. A study by InsideSales.com found that the average salesperson spends 21% of their time on prospecting.
2. According to HubSpot, the cost of generating a single B2B lead can range from $25 to $1,000, depending on the industry and the type of lead generation tactics used
3. Based on this data, time spent on prospecting can cost companies with Large sales teams a significant amount of money

## Goal

To address this problem, the project formulated the problem as a classification task, where the goal was to predict whether an opportunity could be won or not. The project leveraged various machine learning algorithms, including SVM and Random Forest, to build predictive models that could classify leads based on their features. The project also employed various data cleaning, visualization, and manipulation techniques to prepare the data for modeling.

## Overall Approach

The approach of solving the problem involved performing extensive data exploration, cleaning, visualization, and preparation, followed by the implementation of various machine learning algorithms for classification. The models were evaluated using various metrics, and the best performing model was deployed to a production environment to help salespeople sell more efficiently and effectively by providing them with targeted leads and real-time insights.

# Background

Sales prospecting is the process of identifying potential customers or clients for a business and pursuing those opportunities. Traditionally, sales prospecting was done through manual methods such as cold calling, direct mail, and door-to-door sales. These methods were time-consuming, expensive, and often resulted in low-quality leads. Sales teams would spend a significant amount of time and effort on prospecting, which could lead to reduced time and effort on other sales activities.

Over the last 20 years, sales prospecting has evolved significantly with the advent of technology. The rise of the internet and social media has made it easier for sales teams to identify and target potential customers. Email marketing, search engine optimization, and social media advertising have become popular methods for generating leads. Sales teams have also started using customer relationship management (CRM) software to track and manage leads.

However, despite these advances, sales prospecting remains a time-consuming and expensive process. Sales teams still spend a significant amount of time and effort on prospecting, and the quality of leads is often low. This is where AI can make a significant difference. By leveraging AI and machine learning technologies, sales teams can automate routine tasks, provide targeted leads, and offer real-time insights. This can reduce the time and effort spent on prospecting, and help sales teams focus on building relationships with customers and closing deals.

AI can be used in various ways to improve sales prospecting. For example, predictive analytics can be used to identify high-quality leads based on their attributes and behavior. Chatbots can be used to engage with potential customers and answer their questions in real-time. Natural language processing can be used to analyze customer interactions and identify patterns that can be used to improve the sales process. Overall, AI has the potential to revolutionize sales prospecting, making it more efficient, effective, and cost-effective.

# Detailed Approach

## Data Collection

As part of this project, we needed to collect a dataset that included features like sales velocity, cycles, opportunity value, industry, client demographics, and opportunity status. This dataset was necessary for training and testing machine learning models to predict high-quality leads and improve the sales prospecting process. We began by searching for publicly available datasets by Market research firms like Gartner and also reached out to Industry experts like Deloitte to see if they had any relevant data that they could share with us. We were unable to get any data through this channel due to the high level of confidentiality the industries maintain. We then looked at publicly available datasets, like academic research and industry reports and online forums to see if there were any available datasets that met our criteria. We were then able to finalize a dataset from Kaggle which barely everything we were looking for the project

## Exploratory Data Analysis

A crucial step in understanding the data, the data was subjected to various analysis methods. The data exploration was was divided into two parts: Statistical analysis and Visualization

### Statistical Analysis

Our statistical analysis involved several steps. Firstly, we examined the columns of the dataset to understand what kind of information was stored in each column. We then looked at the shape of the data frame, which helped us to understand how much data we were dealing with. We also described the data frame, which gave us a summary of the key statistics for each column. We checked for missing and null values, which helped us to understand the completeness of the dataset. Additionally, we identified any duplicate records, which could skew our results if not handled correctly. Lastly, we listed the numerical and categorical values present in the dataset, which gave us a better understanding of the data types we were working with.

### Visualization

Our visualization analysis was focused on checking the distribution of numerical columns using kernel density plots. This helped us to identify any potential outliers in the data, which could impact the accuracy of our models. We also used boxplots to identify outliers for numerical columns. Additionally, we used count plots for categorical features to understand how balanced the data was. This helped us to identify any class imbalances or biases that might need to be addressed. Finally, we plotted a correlation matrix to identify collinear columns. This helped us to identify any variables that were highly correlated, which could lead to overfitting and reduce the accuracy of our models.

Our data exploration process helped us to gain a deeper understanding of the data we were working with and identify any potential issues that needed to be addressed. By performing statistical analysis and visualization, we were able to get a comprehensive view of the data and prepare it for modeling

## Data Preparation

To prepare the data for modeling, we followed several steps:

**Data imputation:** Imputed outlier values for 3 features (Sales Velocity, Sales Stage Iterations, Opportunity Size (USD)) by replacing them with the 95th percentile to reduce their effect on the model

**Visualization:** Re-visualized the numerical data to check for the presence of any remaining outliers

**Encoding Categorical Variables:** Encoded the feature 'Opportunity Status' which contains values like Won and Loss, to a binary mapping of 0 and 1 to convert them into a format that the model could process and other object data types using a label encoder

**Feature splitting:** Assign the features set to variable X and target feature to y

**Handling Imbalanced Data:** Oversampled the imbalanced target variable using Synthetic Minority Over-sampling Technique (SMOTE) to ensure the model would not be biased towards the majority class

**Splitting Training and Testing data:** Split the data into training and testing sets, ensuring that the split maintained the distribution of the target variable

**Data Scaling:** Scaled the test data feature set to range between 0 and 1 to get each feature on the same scale

**Multicollinearity handling:** Checked for multicollinearity using Variance Inflation Factor mechanism to ensure that the model would not be affected by the high correlation between predictor variables

## Modeling

To classify whether an opportunity is a win or loss, we used two popular machine learning algorithms: Support Vector Machines (SVMs) and Random Forest Classifier.

We began by importing the necessary libraries for both algorithms, including scikit-learn for model building and pandas for data manipulation. We then initialized the models, specifying the hyperparameters such as the kernel function and regularization parameter for SVM, and the number of estimators, maximum depth of the tree, and minimum number of samples required to split an internal node for Random Forest Classifier.

We fit both models with the training data, which involved the encoded feature variables and target variable

## Model Evaluation

After fitting, we used the models to make predictions on the test data. We evaluated the performance of both models using accuracy, precision, recall, and F1-score metrics as seen below

| Evaluation Metric | Support Vector Machine | Logistic Regression | Random Forest Classifier |
|---|---|---|---|
| Test Accuracy | 0.774801 | 0.689754 | 0.835536 |
| Precision | 0.790440 | 0.686642 | 0.836486 |
| Recall | 0.748097 | 0.698499 | 0.834271 |
| F1-score | 0.768686 | 0.692520 | 0.835377 |

## Hyper Parameter Tuning

In order to optimize the performance of our two machine learning algorithms – Support Vector Machines (SVMs) and Random Forest Classifier – we performed hyperparameter tuning using Grid Search. Hyperparameter tuning allowed us to find the best combination of hyperparameters for each model to yield the highest possible accuracy and F1-score.

For the SVM, we tuned the kernel function, regularization parameter (C), and the gamma parameter. For the Random Forest Classifier, we optimized the number of estimators, maximum depth of the tree, minimum number of samples required to split an internal node, and other hyperparameters.

After tuning the hyperparameters using Grid Search, we retrained the models and evaluated their performance using the same accuracy, precision, recall, and F1-score metrics as before. The results showed an improvement in model performance across both algorithms

## Deployment

To deploy the best performing model, the Random Forest Classifier, we used a combination of Python libraries and web technologies. We first serialized the scaler, label encoder, and the model using the pickle library to save the pre-processing and model configurations for future use. We then built a lightweight web application with Flask, which served as an interface for users to input relevant data and receive predictions on whether a sales opportunity would result in a win or a loss. We designed a clean and user-friendly interface using HTML and CSS for sales team members to input required data.

We integrated the pre-trained model into our Flask application by loading the saved pickle files, allowing us to preprocess user input and make predictions without retraining the model. Finally, we deployed the Flask web application on a web server, making it accessible to the sales team. This practical tool helped them prioritize high-quality leads and invest their time more effectively, leading to increased sales and overall business growth.

# Results Explanation

The results section emphasizes the effectiveness of our chosen model, the Random Forest Classifier, in predicting the outcome of sales opportunities as either a win or a loss. This model demonstrated a strong ability to accurately classify opportunities, providing valuable insights for the sales team. By using this tool, they can better prioritize their efforts and focus on high-potential leads.

The visual representation of the model's performance showcases its strengths and weaknesses, allowing team members to adjust their approach accordingly. Although the model does not provide direct insights into the most influential factors contributing to the outcome, its accurate predictions can still guide the sales team in their decision-making process.

Despite being a sophisticated tool, the Random Forest Classifier is user-friendly and accessible to sales team members. The model's interpretability builds trust and promotes its adoption among the team, as they can understand the rationale behind the predictions.

In conclusion, the results obtained from the Random Forest Classifier demonstrate its ability to effectively predict the outcome of sales opportunities as either a win or a loss. Integrating this model into the sales team's daily workflow enables them to make data-driven decisions, prioritize high-quality leads, and ultimately achieve better results and business growth.

# Use Case

A potential practical application of the developed methodology could be in technology firms that offer a diverse range of products, such as software services, subscriptions, and hardware devices. One such example is VmWare, which boasts of a rapidly expanding customer base of over 500,000. With thousands of leads to process daily, the task of analyzing each opportunity's potential can be overwhelming for sales representatives. Our project presents a solution that can enable effective sales planning by optimizing the process of identifying leads that are worth pursuing, allowing sales representatives to focus on selling rather than spending valuable time analyzing each opportunity. This approach can improve the efficiency of the sales process, resulting in more profitable outcomes for the firm.

# Conclusions

## Summary

Overall, an AI ML solution for helping salespeople sell more by spending less time prospecting has the potential to transform the way sales teams operate, by providing them with a more efficient and effective way to identify and engage with customers. By leveraging advanced technologies to automate routine tasks and provide real-time insights, sales teams can focus on building relationships and closing deals, ultimately driving business growth and success

## Limitations

One limitation of the developed algorithm is its restriction to the Information Technology sector in India, as it was trained on this specific dataset. This may cause the algorithm to behave differently when applied to data from other industries or countries. It is important to note that the algorithm's performance on other datasets needs to be evaluated before implementing it in different scenarios. Further research and development of the algorithm may be necessary to improve its adaptability to different contexts.

## Future Work

Sales Opportunity Classification is just one of the pieces of Sales Planning. Future work in the same area can be expanded to:

1. Sales forecasting: AI and ML can be used to predict sales revenue, quantity, and trends based on historical data, market trends, and other factors.
2. Lead generation and prospecting: AI and ML can be used to identify potential customers and generate leads based on demographic data, online behavior, and other factors.
3. Sales performance management: AI and ML can be used to analyze sales data and performance metrics to identify areas for improvement and optimize sales strategies.
4. Sales territory optimization: AI and ML can be used to analyze sales data and customer data to optimize sales territories and maximize revenue.
5. Customer segmentation and targeting: AI and ML can be used to analyze customer data and behavior to segment customers and create targeted marketing campaigns

# References


https://blog.hubspot.com/sales/sales-statistics

https://www.forbes.com/sites/kenkrogue/2018/02/15/why-sales-reps-spend-so-little-time-selling/?sh=4752daa61051

https://www.linkedin.com/pulse/why-sales-reps-only-spending-352-time-selling-gabe-larsen/

https://blog.thebrevetgroup.com/21-mind-blowing-sales-stats

https://blog.thecenterforsalesstrategy.com/6-keys-to-know-which-prospects-are-worth-the-dip

https://www.prosalesconnection.com/blog/how-much-time-should-your-sales-team-devote-to-prospecting

https://www.paceproductivity.com/single-post/2017/02/09/how-sales-reps-spend-their-time