<u>INTRODUCTION:</u>

<u>OVERVIEW:</u>

The project "Wastewhiz: Data Science-Enabled Marketing Efficiency Enhancement" aims to leverage Data Science techniques to enhance the efficiency of marketing efforts for an organization providing a hiring assessment platform. The objective is to develop a sophisticated Machine Learning model that can predict the probability percentage of marketing leads converting into product purchasers. By analyzing the provided dataset, which contains relevant information, the model will generate predictions based on key fields. This predictive model will enable the organization to optimize its marketing strategy, reduce yearly marketing expenses, and improve overall conversion rates. The skills required for this project include Exploratory Data Analysis, proficiency in MS Excel, and experience with IBM Cognos Analytics.

<u>PURPOSE:</u>

The purpose of the "Wastewhiz: Data Science-Enabled Marketing Efficiency Enhancement" project is to revolutionize the organization's marketing approach by harnessing the power of Data Science and Machine Learning. By developing a predictive model that estimates the probability of marketing leads converting into product purchasers, this project aims to achieve the following:

◎ Optimized Marketing Strategy: The predictive model will enable the organization to focus its marketing efforts on leads with the highest probability of converting. This targeted approach will lead to a more efficient allocation of resources and a reduction in marketing expenditures.

◎ Cost Reduction: By identifying leads that are less likely to convert, the organization can avoid unnecessary marketing expenses on individuals or segments that may not yield significant returns. This will contribute to substantial cost savings over time.

◎ Improved Conversion Rates: Through accurate predictions, the organization can tailor its marketing messages and strategies to align with the preferences and characteristics of potential customers. This customization will enhance the

likelihood of successful conversions.

◎ Enhanced Decision-Making: The insights gained from the predictive model's analysis of the dataset will provide valuable information about the factors influencing marketing success. This knowledge can guide informed decision-making in designing future marketing campaigns.

◎ Competitive Edge: Implementing a data-driven approach will position the organization as an innovator in the industry. The ability to efficiently predict and target potential customers sets the stage for outperforming competitors and staying ahead in the market.

In essence, this project seeks to transform the organization's marketing efforts from a traditional and resource-intensive approach to a data-driven, precision-targeted strategy. By accurately estimating the likelihood of marketing success, the organization can streamline its operations, reduce costs, and achieve higher conversion rates, ultimately driving business growth and success.

LITERATURE SURVEY:

Existing Problem:

WasteWhiz is a waste management company that aims to optimize its marketing efforts through data science techniques. The company operates in multiple regions, offering management solutions to various industries and businesses aiming to predict their success probabilities with respect to certain major factors.

However, they face challenges in efficiently targeting potential customers and optimizing their marketing strategies to maximize conversions and minimize costs and some challenges in predicting the success probabilities using conventional methods. Also they have not considered many key factors for predicting the success probability which may play a major role contributing towards the success probability like the level of meeting nor the correlation between the industry and the geographical location at that time.

Proposed Solution:

Using IBM Watson Studio we have developed a data science-enabled tool, WasteWhiz, that enhances the company's marketing efficiency by leveraging data analysis, predictive modelling, and customer segmentation techniques. WasteWhiz will provide actionable insights improve customer targeting, and enhance overall conversion rates. We have used necessary data science techniques on our retrieved dataset, have analyzed thoroughly the data and we have built a data science model using random forest algorithm with high accuracy that could predict the success probabilities. We have used Django framework for our website WasteWhiz which provides the success probability and we have considered many key factors and small factors that could play a role in determining the success probability of the corresponding product.

THEORETICAL ANALYSIS:

Hardware / Software designing:

Hardware:

- 64-bit Windows 8 and above

- 8 GB RAM

Software:

- Python 3.8 and above

- Django version 4.2

- IBM Watson Studio

EXPERIMENTAL INVESTIGATIONS:

DATA PREPROCESSING:

- Some unimportant columns like Contact_no, POC_name, Deal_title, Lead_name, etc. were removed from the dataframe as they do not impact the marketing efficiency much.

- The Deal_value and Weighted_amount columns were converted to integers after removing the $ symbol.

- The columns Industry, Pitch, Lead_revenue, Fund_category, Geography, Designation, Hiring_candidate_role, Lead_source, Last_lead_update and Resource were converted from categorical variables to a number using get_dummies().

OUTLIER TREATEMENT:

An interesting approach was taken for treating outliers in the Deal_value and Weighted_amount column. Here the standard deviation times a certain threshold is multiplied and is added and subtracted to the mean to arrive at appropriate intervals.

ANALYSIS OF DATA:

- For Analysing the data consider the following histogram that has been constructed from the train dataset after removing outliers and cleansing the data:
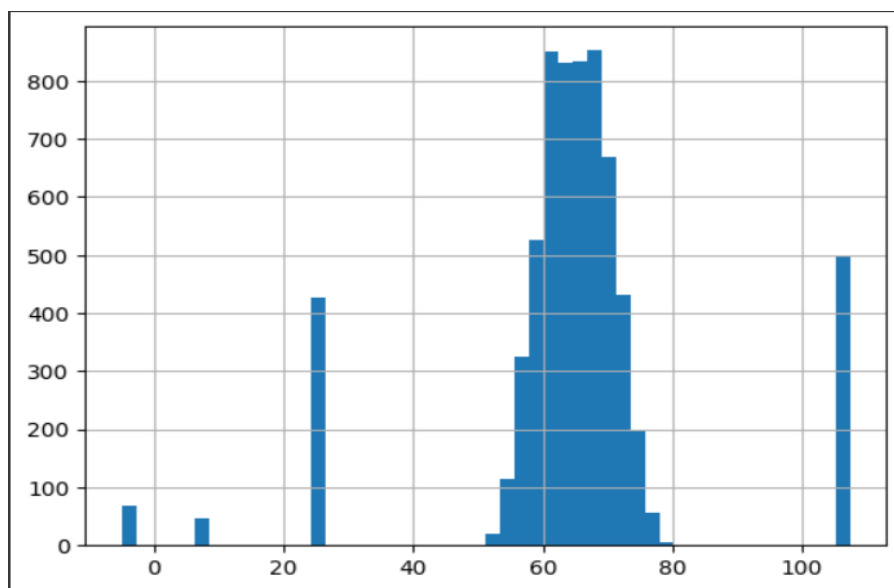
- This is a step-filled histogram of Success_probability which shows that only those rows with a success probability between around 50% to 80% follow a normal distribution.
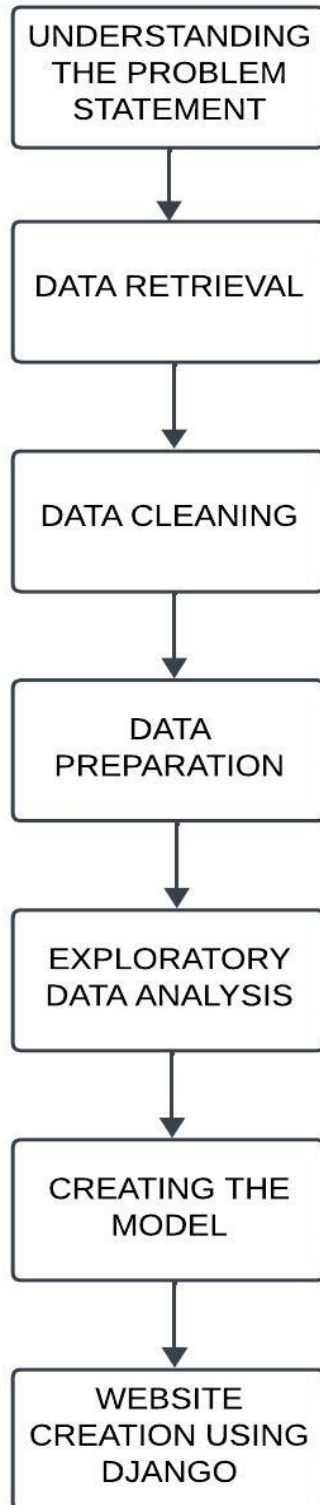
MODEL DEVELOPMENT:

- The Random Forest model which a supervised learning model. Here we use it for solving a regression problem.

- Random Forest is an ensemble method i.e., it combines the predictions from other models to give an accurate result.

- A novel idea implemented here is that of considering the weights of the two types of data i.e., those following normal distribution and those that do not. Here it is implemented using np.where().
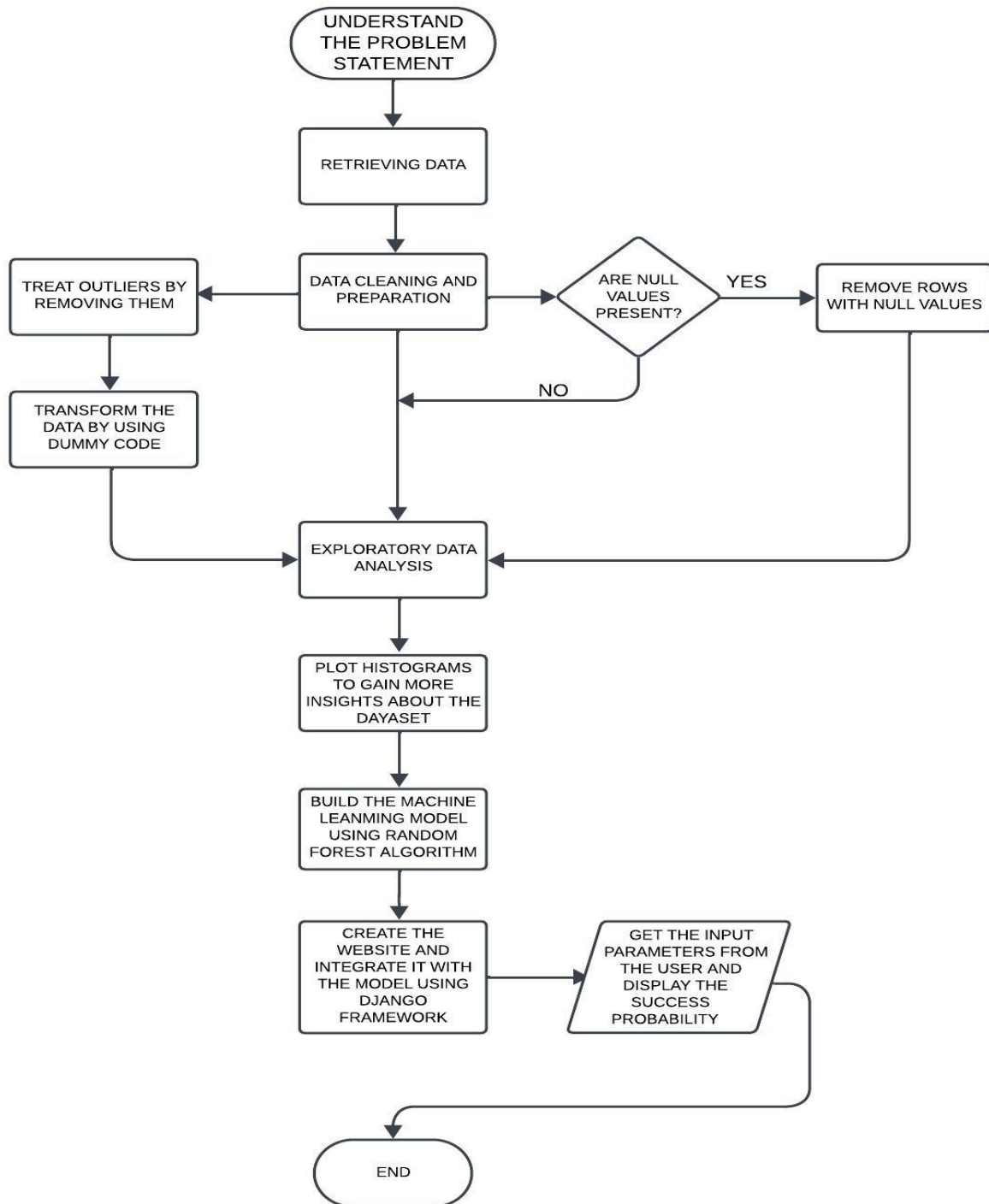
DESIGNING THE WEBPAGE:

- The entire project was converted into a webpage using HTML and Django.

- An existing template was used, and the required fields were added as text boxes.

- The ML model is run on the background and the result is evaluated according to the data entered.

BLOCK DIAGRAM:

```
┌─────────────────────┐
│    UNDERSTANDING    │
│     THE PROBLEM     │
│      STATEMENT      │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    DATA RETRIEVAL   │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    DATA CLEANING    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│        DATA         │
│     PREPARATION     │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    EXPLORATORY      │
│   DATA ANALYSIS     │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│   CREATING THE      │
│       MODEL         │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│      WEBSITE        │
│  CREATION USING     │
│      DJANGO         │
└─────────────────────┘
```

FLOW CHART:

<u>APPLICATIONS:</u>

1. **<u>Marketing Campaign Optimization:</u>**

WasteWhiz can analyze with pre-available data
and market trends to predict the success probabilities accurately.

2. **<u>Predictive lead scoring:</u>**

The tool uses proper data science techniques to predict the success probabilities taking into consideration a number of important factors that are in most cases neglected.

3. **<u>Personalization of Content:</u>**

WasteWhiz can provide personalized
recommendations or solutions about the product depending on its success probability analyzing on the key factors to be taken care of for increasing the success probability.

4. **<u>Market Trend Analysis:</u>**

By analyzing broader market trends and preferences from its previous data WasteWhiz can provide insights into increasing the rate of success according to the needs and emerging opportunities and threats.

<u>CONCLUSION:</u>

In culmination, the "Wastewhiz: Data Science-Enabled Marketing Efficiency Enhancement" project represents a significant leap forward in the organization's marketing strategies, made possible by the seamless integration of Data Science, Machine Learning, and web development. The project's journey can be summarized as follows:

1. **Data Analysis and Model Training:** The project began with an in-depth exploration of the provided dataset. Through meticulous Exploratory Data Analysis, the team identified key features that impact marketing success. The resulting insights informed the training of a sophisticated Machine Learning model.

2. **Model Storage:** The trained model's capabilities were encapsulated in a Python script, `model.py`, which upon execution, generated a model file named `ml_model.sav`. This file, the culmination of data-driven insights and machine learning techniques, held the predictive power to estimate the probability of marketing success.

3. **Website Development:** Leveraging the Django framework, the project translated the predictive model into an accessible and user-friendly web application. The resulting website provides a seamless interface for users to input relevant lead features.

4. **Real-Time Prediction:** By utilizing the stored `ml_model.sav` file, the website dynamically creates the model within seconds, leveraging the pre-trained machine learning model. Upon user input, the website processes the data through the model, generating accurate predictions of marketing success probability.

5. **Streamlined Decision-Making:** Marketing teams now have an invaluable tool at their disposal. The website's immediate output of predicted success probabilities empowers teams to make agile and informed decisions, optimizing resource allocation and maximizing conversion rates.

In essence, the project's outcome is a testament to the synergy of data, technology, and innovation. By crafting a robust Machine Learning model and embedding it within a Django-based website, the organization has revolutionized its marketing approach. The streamlined process from model training to real-time prediction empowers the organization to forecast marketing success with precision and confidence. This project stands as a beacon of how data science and technology can collaboratively propel businesses into a new era of efficiency and strategic decision-making.

FUTURE SCOPE:

Advanced AI and Machine Learning Techniques:
Incorporate more advanced AI and machine learning techniques as they develop, such as deep learning, reinforcement learning, and natural language processing. This could enhance the accuracy of predictive models and insights.

Real-time Analytics:
Develop capabilities for real-time data analysis, allowing WasteWhiz to provide immediate insights and recommendations based on up-to-the-minute data.

Multichannel Integration:
Extend WasteWhiz to encompass a wider range of marketing channels, such as social media, email, mobile apps, and more. This would provide a holistic view of customer interactions across various touchpoints.

Intelligent Content Generation:
Incorporate AI-powered content generation for marketing materials, such as personalized emails, social media posts, and product recommendations.

Geographic and Cultural Considerations:
Tailor marketing strategies based on geographic and cultural nuances, ensuring global relevance and effectiveness.

Bench marking and Comparative Analysis:
Introduce features that enable the company to compare its marketing performance against industry benchmarks and competitors.

<u>ADVANTAGES:</u>

- This solution uses only relevant data columns/fields thereby reducing the time of computation and increasing the accuracy of the solution.

- Random Forest is an accurate and robust model. It provides higher accuracy compared to individual decision trees due to reduced over fitting.

- The given data was not directly used in the model. Weights were used to handle the different types of data.

- The project was converted to a website making user friendly and interactive.

<u>DISADVANTAGES:</u>

- The interpretation of the results is important. Since the solution is based on statistics and data science, there is always a margin of error in the accuracy of the results.

- The solution does not use the location column but only uses Geography. The same solution can be extended using location to give better results.

- This process is dependent on the quality of the data used. If the data is of poorer quality, then the results are not accurate.

<u>BIBLIOGRAPHY:</u>

1. https://www.geeksforgeeks.org/python-pandas-get_dummies-method

2. https://stats.oarc.ucla.edu/spss/faq/coding-systems-for-categorical-variables-in-regression-analysis-2/#:~:text=Categorical%20variables%20require%20special%20attention,entered%20into%20the%20regression%20model

3. https://www.kaggle.com/datasets/bhavikjain/reduce-marketing-waste-hackerearth-ml-challenge

<u>APPENDIX:</u>

Refer to the Github repo for the project source code.