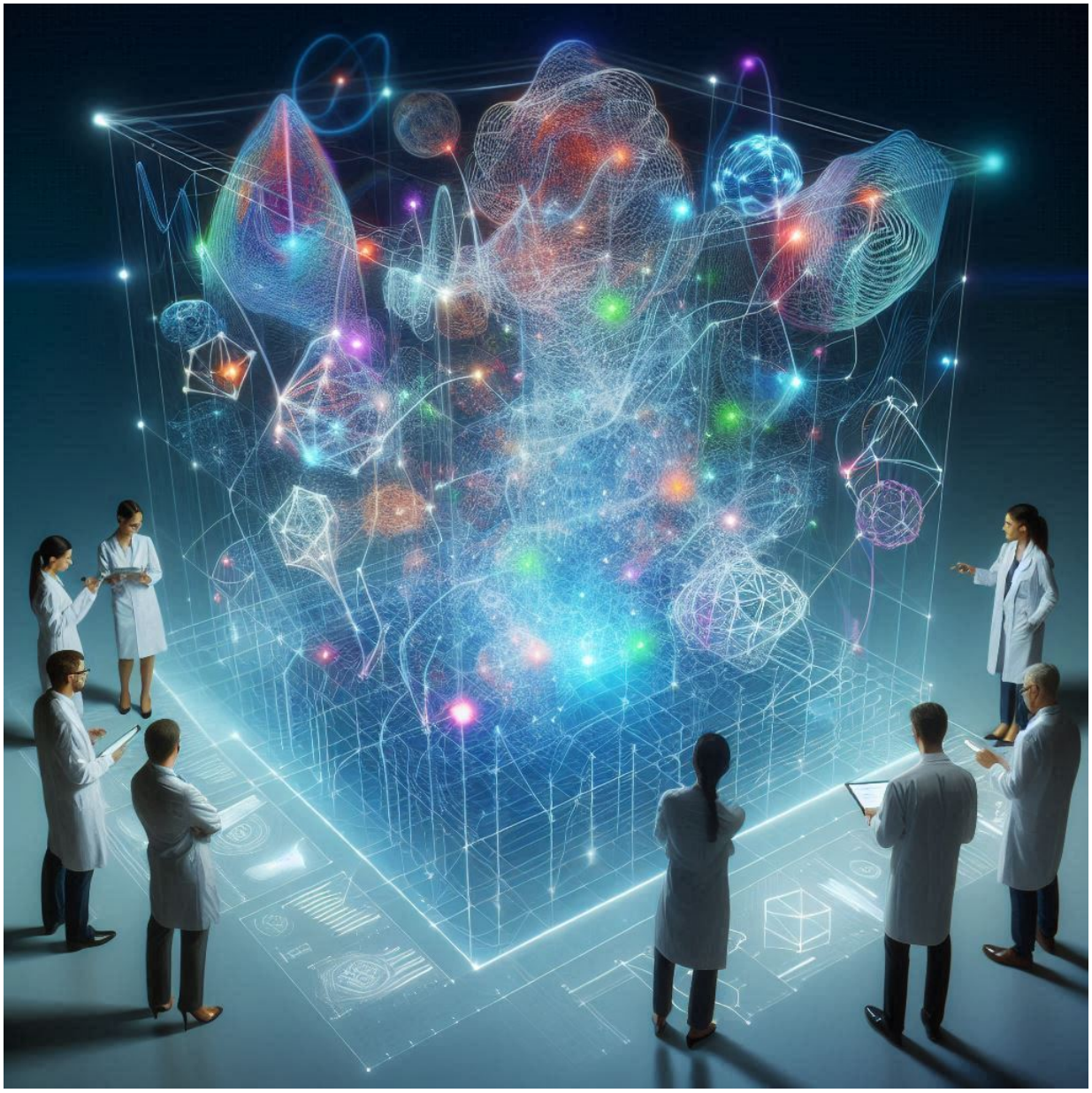
# Propensify (Propensity Model to identify how likely certain target groups customers respond to the marketing campaign)



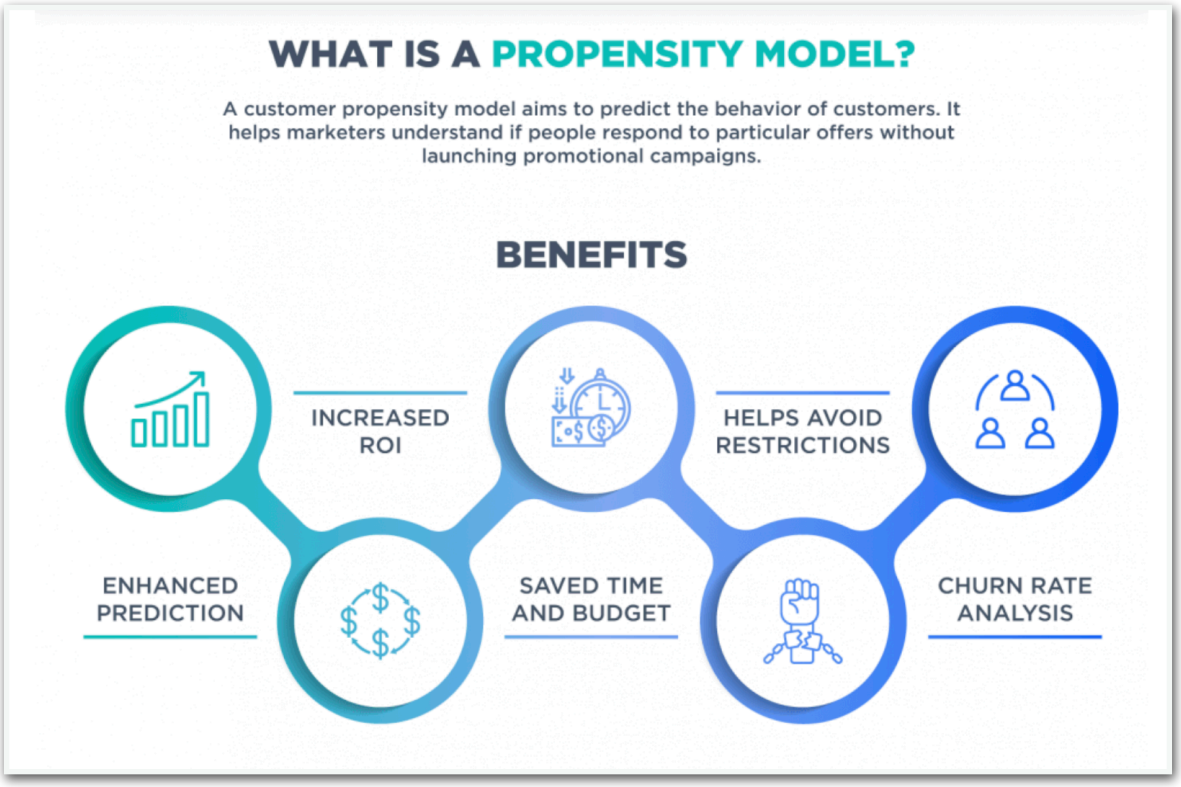


Raju kumar

# Project 1: Propensity Model for Predictive Marketing in Insurance

## Context

In the competitive landscape of insurance marketing, understanding customer behavior and predicting their actions is crucial. Insurance companies face the challenge of efficiently targeting potential customers to maximize marketing effectiveness and minimize costs. The idea behind predictive marketing is to leverage data analytics to identify customers who are most likely to purchase insurance policies, thereby optimizing marketing efforts and improving operational efficiency.




## Objective

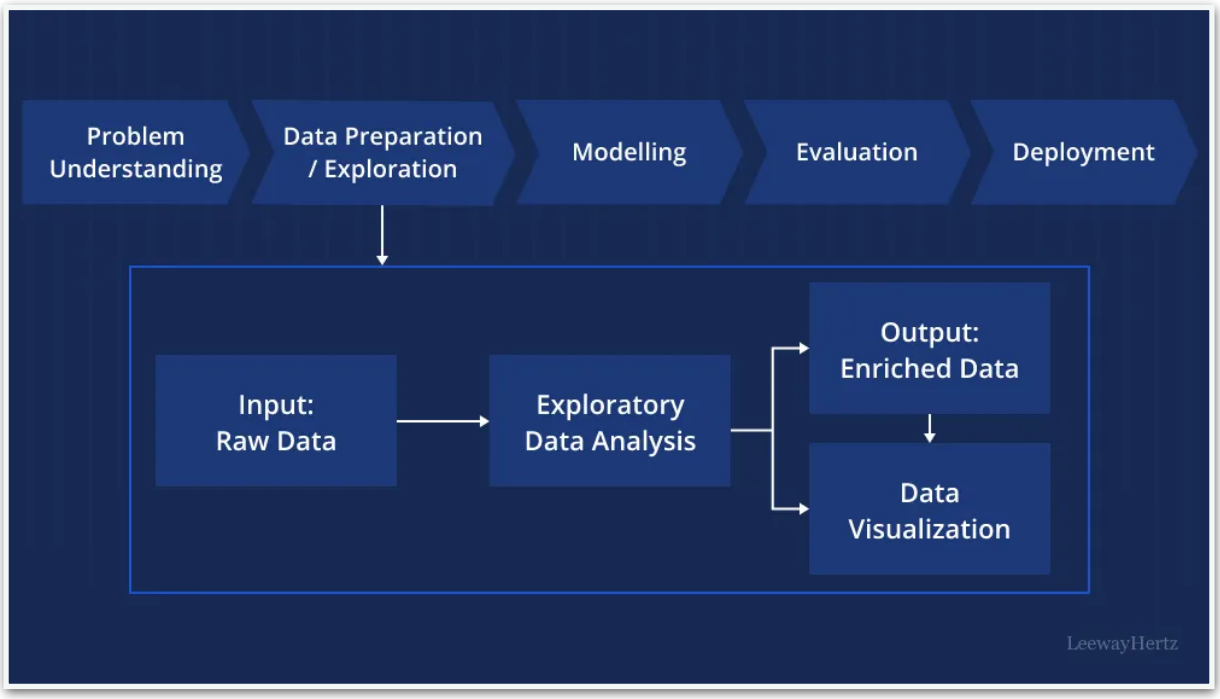
The objective of this propensity model project is to predict the likelihood of individuals purchasing an insurance policy by identifying irregularities and patterns in the collected data. This allows organizations to take targeted marketing actions, optimize their marketing strategies, and improve overall business performance.

## The Data

An insurance company provided a file, train.csv, as a historical dataset, and test.csv containing the list of people whom they want to target with their marketing. From this list of people, one needs to determine the yes/no decision about whether to market to them by using the relevant columns provided in the datasets.

| Type | Name | Description |
|---|---|---|
| Input Variables | custAge | The age of the customer (in years) |
| Input Variables | profession | Type of job |
| Input Variables | marital | Marital status |
| Input Variables | schooling | Education level |
| Input Variables | default | Has a previous defaulted account? |
| Input Variables | housing | Has a housing loan? |
| Input Variables | loan | Has a personal loan? |
| Input Variables | contact | Preferred contact type |
| Input Variables | month | Last contact month |
| Input Variables | day_of_weel | Last contact day of the week |
| Input Variables | campaign | Number of times the customer was contacted |
| Input Variables | pdays | Number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted) |
| Input Variables | previous | Number of contacts performed before this campaign and for this client |
| Input Variables | poutcome | Outcome of the previous marketing campaign |
| Input Variables | emp.var.rate | Employment variation rate - quarterly indicator |
| Input Variables | cons.price.id | Consumer price index - monthly indicator |
| Input Variables | cons.conf.idx | Consumer confidence index - monthly indicator |
| Input Variables | euribor3m | Euribor 3 month rate - daily indicator |
| Input Variables | nr.employed | Number of employees - quarterly indicator |
| Input Variables | pmonths | Number of months that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted) |
| Input Variables | pastEmail | Number of previous emails sent to this client |
| Target Variables | responded | Did the customer respond to the marketing campaign and purchase a policy? |

## EXPLORATORY DATA ANALYSIS (EDA)




### Purpose

EDA helps us gain insights into the dataset and understand its characteristics.

**Actions and Motive**

1. **Data Patterns and Relationships**

- **Examine statistical summaries (mean, median, standard deviation) for each feature:**
  - **Motive:** To understand the distribution and central tendency of the features which can help in identifying patterns or anomalies in the data.
- **Identify correlations between features (using correlation matrices or scatter plots):**
  - **Motive:** To detect relationships between variables, which can inform feature selection and engineering processes.

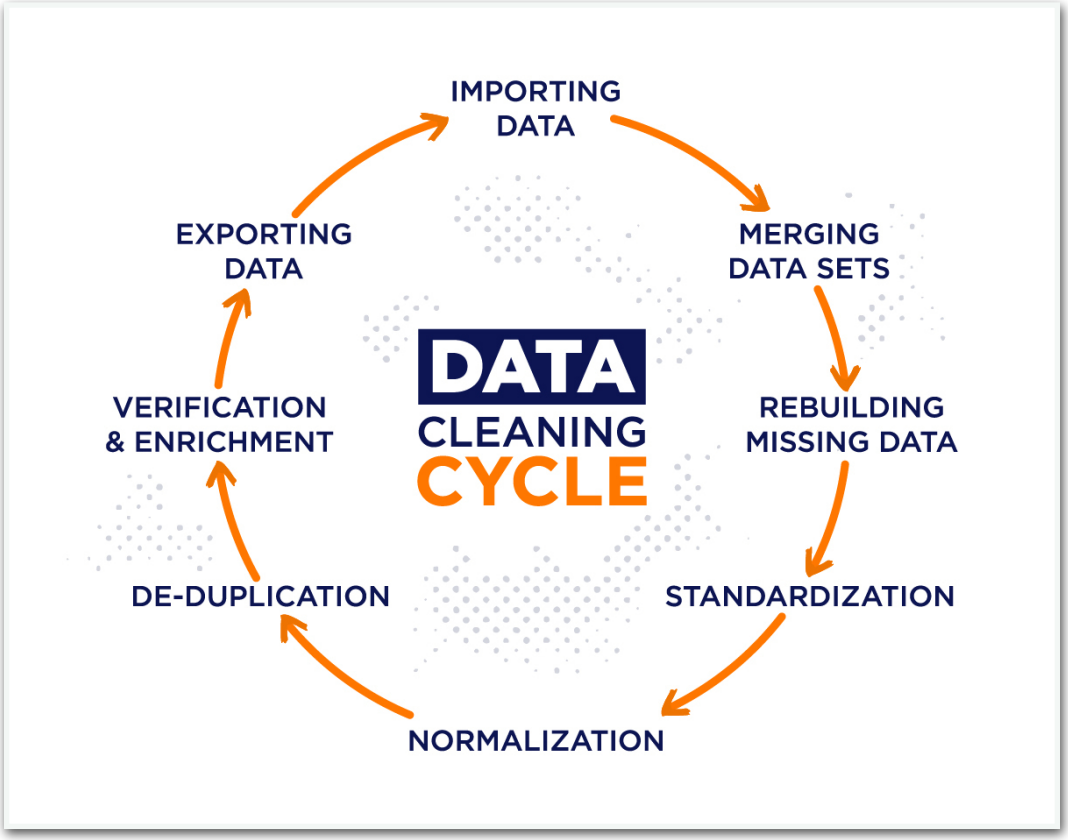
2. **Handling Missing Values and Outliers**

- **Check for missing data and decide on an imputation strategy (mean, median, or other methods):**
  - **Motive:** To ensure completeness of the dataset as missing values can skew the model performance.
- **Detect outliers using methods like the Interquartile Range (IQR) and decide whether to remove or transform them:**
  - **Motive:** To manage extreme values which can disproportionately affect model training and accuracy.

3. **Visualization**

- **Create histograms, box plots, and scatter plots to visualize feature distributions and relationships:**
  - **Motive:** To get a visual understanding of the data, identifying skewness, outliers, and potential transformations needed.
- **Explore time series patterns if applicable (e.g., customer behavior over time):**
  - **Motive:** To understand trends and seasonal patterns that could influence customer behavior.

# DATA CLEANING AND STANDARDIZATION




**Purpose**

Ensure data quality and consistency.

**Actions and Motive**

1. **Standardization**

- **Convert features to a common scale (e.g., z-score normalization) to avoid bias:**
  - **Motive:** To ensure that features contribute equally to the model, avoiding bias towards features with larger scales.

2. **Handling Missing Values**

- **Impute missing data (mean, median, or other methods) based on context:**
  - **Motive:** To fill in missing information, maintaining the integrity of the dataset for effective modeling.
- **Consider domain knowledge when making imputation decisions:**
  - **Motive:** To make informed decisions that align with business logic and improve imputation accuracy.

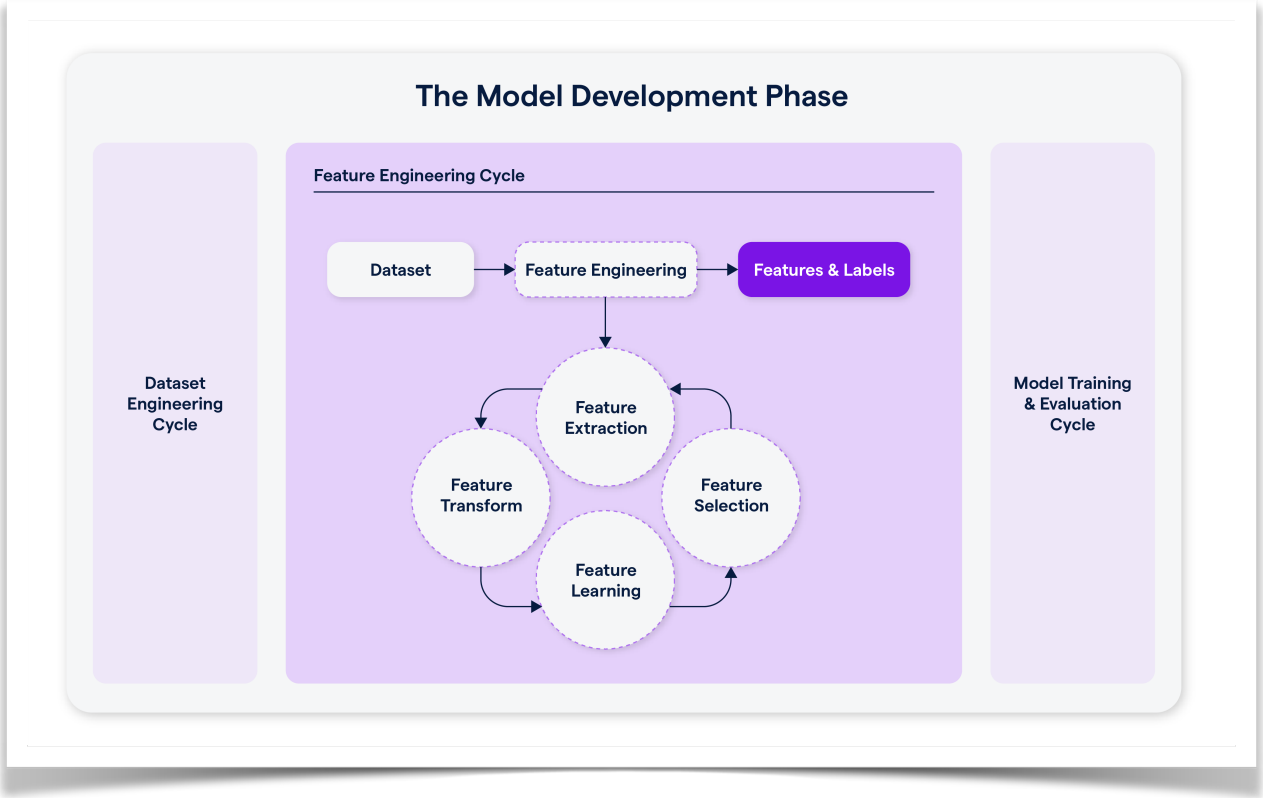
3. **Outlier Treatment**

- **Identify extreme values using IQR or other methods:**

- **Motive:** To detect values that deviate significantly from other observations.

- **Decide whether to remove outliers or transform them (e.g., winsorization):**
  - **Motive:** To mitigate the impact of outliers on model training, which can improve model performance and stability.

## FEATURE ENGINEERING




**Purpose**

Enhance model performance by creating relevant features.

**Techniques and Motive**

1. **Aggregation**

- **Combine related features (e.g., sum, average, or count of specific columns):**
  - **Motive:** To create new features that encapsulate the combined effect of multiple variables, enriching the dataset.

2. **Transformation**

- **Apply mathematical functions (e.g., log, square root) to existing features:**

- **Motive:** To stabilize variance and normalize skewed distributions, making patterns more discernible.

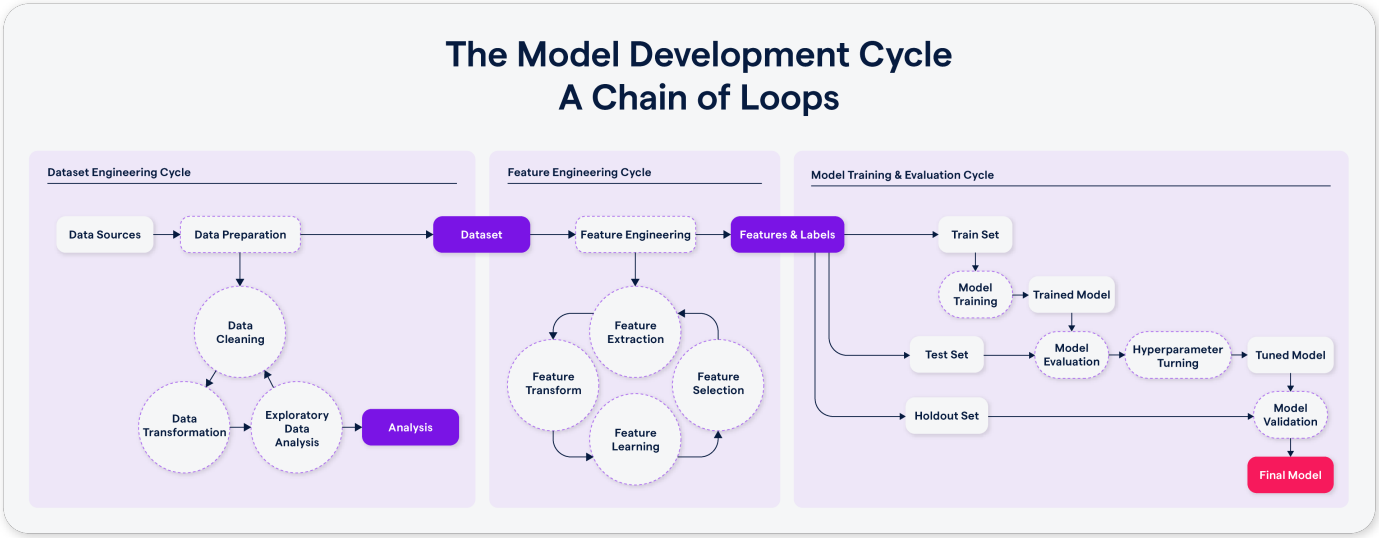
3. **Interaction Terms**

- **Create new features by combining existing ones (e.g., product of two features):**
  - **Motive:** To capture relationships between variables that are not apparent when considered individually.

4. **Domain-Specific Features**

- **Engineer features based on domain knowledge (e.g., time-based features):**
  - **Motive:** To leverage industry-specific insights, improving the relevance and accuracy of the predictive model.

## MODEL SELECTION AND TRAINING




### Purpose

Choose an appropriate machine learning model and train it.

### Actions and Motive

1. **Model Choice**

- **Consider the problem type (classification for propensity modeling) and data size:**

- **Motive:** To select a model that is well-suited for the binary classification task and scalable to the dataset size.

- **Common choices include Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, Support Vector Classifier, K-Nearest Neighbors Classifier (KNN), XGBoost Classifier, and Neural Network Classifier:**
  - **Motive:** To explore a variety of algorithms to find the most effective one for the specific problem context.

2. **Train-Test Split**

- **Divide data into training and testing subsets (e.g., 80% train, 20% test):**
  - **Motive:** To evaluate the model's performance on unseen data, ensuring its generalization capability.

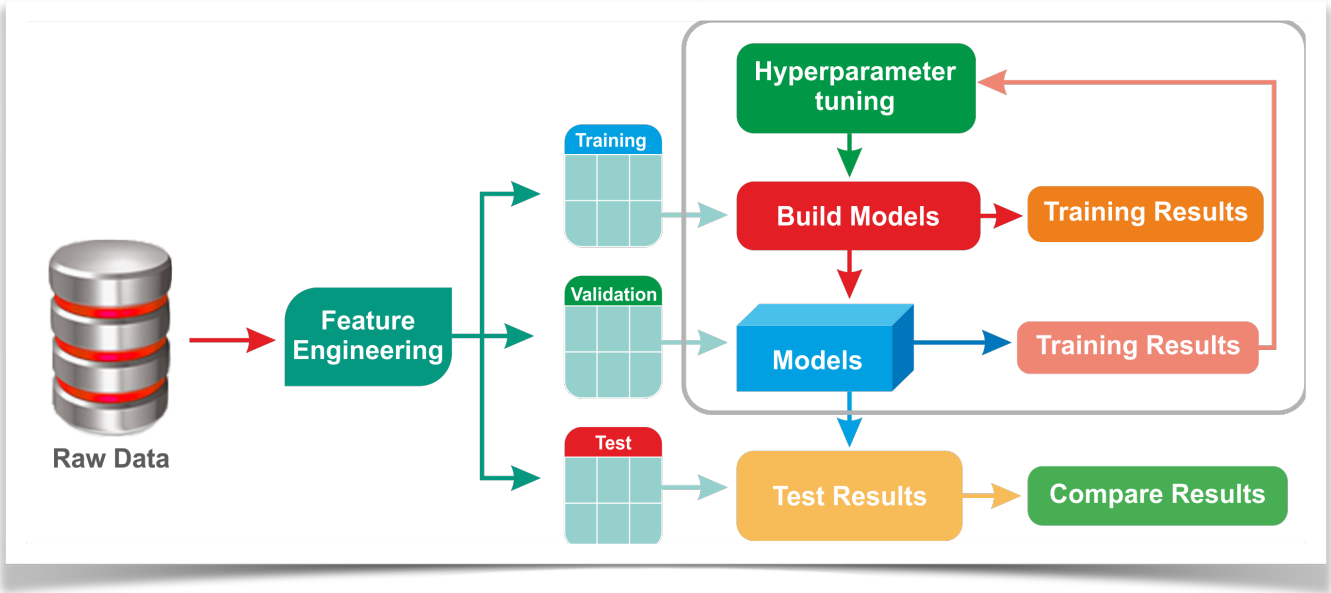
3. **Model Training**

- **Fit the chosen model on the training data:**
  - **Motive:** To enable the model to learn patterns from the data, forming the basis for making predictions.

4. **Evaluation Metrics**

- **Select appropriate metrics (accuracy, precision, recall, F1-score, ROC-AUC) for model evaluation:**
  - **Motive:** To assess the model's performance comprehensively, ensuring it meets the business objectives.

## HYPERPARAMETER TUNING AND MODEL VALIDATION



**Purpose**

Optimize model performance and assess generalization.

**Actions and Motive**

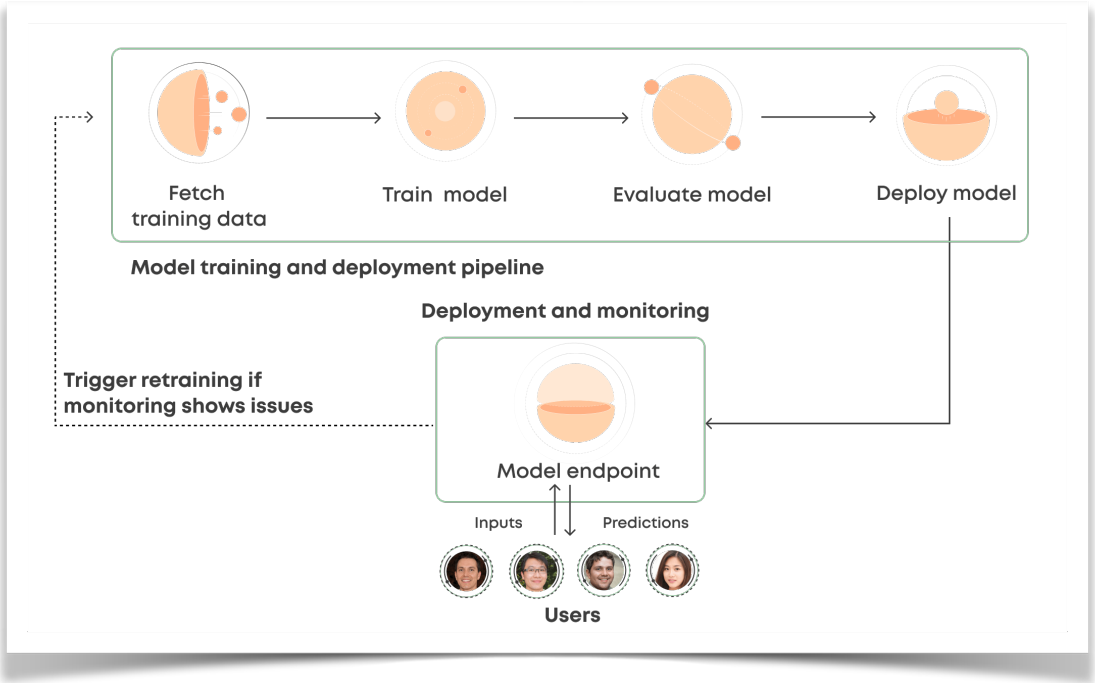
1. **Hyperparameter Tuning**

- **Use techniques like grid search or random search to find optimal hyperparameters:**
  - **Motive:** To refine the model settings for enhanced accuracy and performance.
- **Tune parameters such as learning rate, max depth, or regularization strength:**
  - **Motive:** To improve model efficiency and prevent overfitting or underfitting.

2. **Cross-Validation**

- **Perform k-fold cross-validation to estimate model performance on unseen data:**
  - **Motive:** To ensure the model's robustness and reliability by evaluating it across multiple data splits.
- **Address overfitting by monitoring training and validation scores:**
  - **Motive:** To ensure the model generalizes well to new data, balancing bias and variance.

## MODEL DEPLOYMENT PLAN



Prepare the trained model for production use.

**Actions and Motive**

1. **Integration**

- **Integrate the model into the existing system (e.g., API, web service):**
  - **Motive:** To deploy the model in a live environment where it can generate real-time predictions.

2. **Monitoring**

- **Continuously monitor model performance and retrain if necessary:**
  - **Motive:** To maintain model accuracy and adapt to any changes in data patterns over time.

3. **Scalability**

- **Ensure the model can handle real-time data and scale as needed:**
  - **Motive:** To support growing data volumes and maintain performance under increased load.

## FUTURE DIRECTIONS

As we chart the course ahead, the evolution of the propensity model gains momentum.

- **Real-time Vigilance**

- **Our compass points toward streaming data integration:**
  - **Motive:** To provide timely alerts and proactive marketing actions, preventing missed opportunities before they unfold.
- **Ensemble Exploration**

- **We delve into the art of ensemble methods:**
  - **Motive:** To combine models and unlock robust predictions, enhancing model accuracy and stability.
- **Agile Adaptation**

- **The heartbeat of the propensity model lies in continuous improvement:**
  - **Motive:** To ensure regular model updates, fueled by fresh data insights, propelling us toward efficiency and excellence.

## SUMMARY

The Propensity Model project has effectively built an automated system for predictive marketing in the insurance industry. With the astute application of machine learning methodologies and rigorous preprocessing of the data, the model has achieved a high level of accuracy in predicting customer behavior. This is not only a milestone achieved but also forms a foundation for further optimization and industrial application in various industries.

## Source Code

The source code used to create the pipeline can be found in the attached files.

Thank you for the opportunity to work on the Propensity Model project. For any further inquiries or assistance, please feel free to reach out.

Sincerely,

RAJU KUMAR