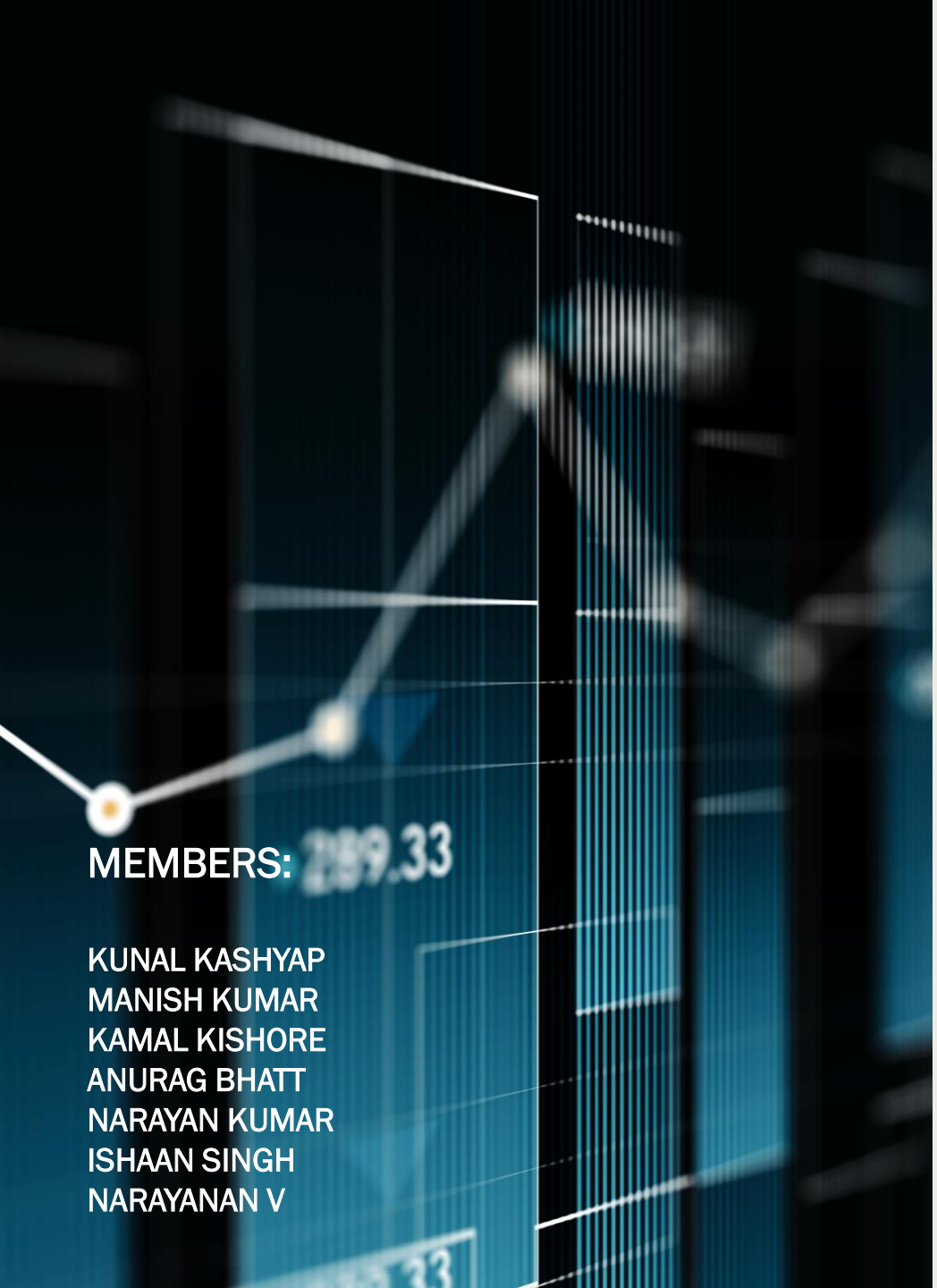


MEMBERS:

KUNAL KASHYAP
MANISH KUMAR
KAMAL KISHORE
ANURAG BHATT
NARAYAN KUMAR
ISHAAN SINGH
NARAYANAN V

LEAD OPTIMIZATION MODEL

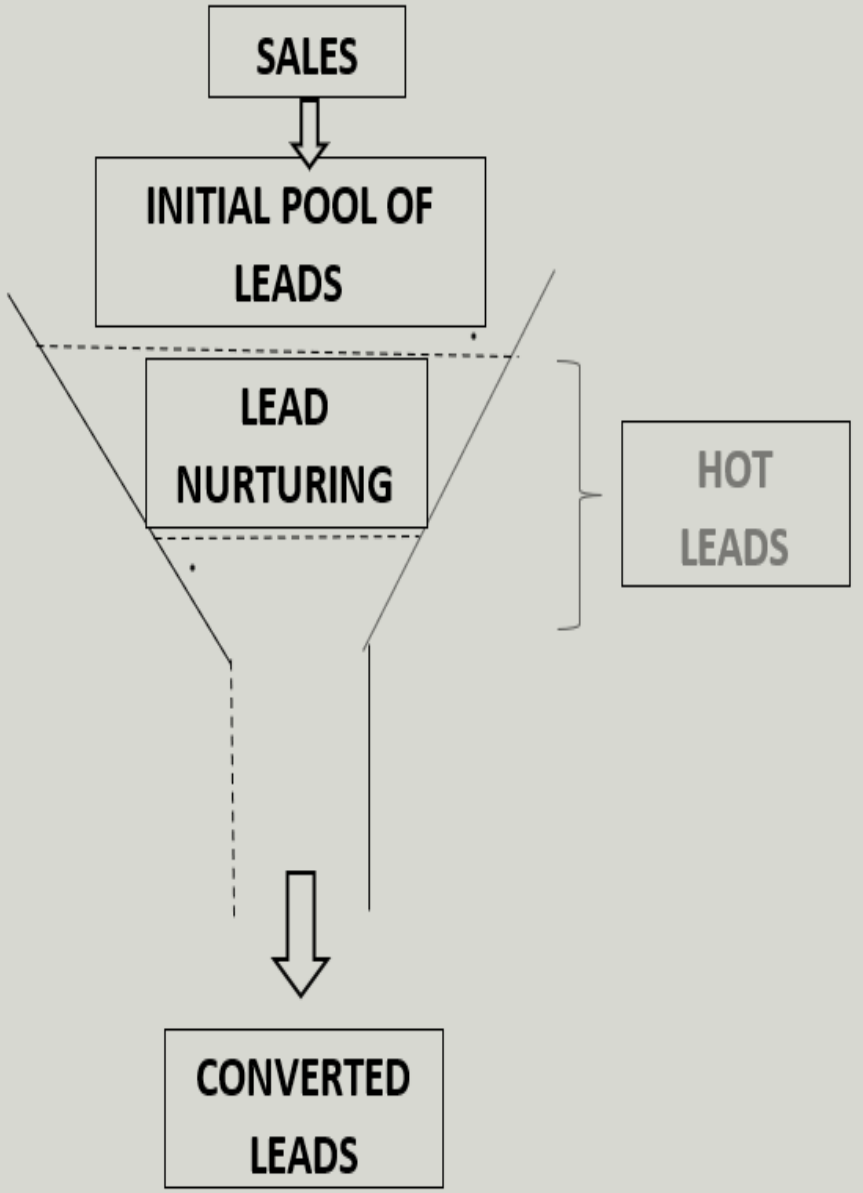
Machine learning - driven Approach for predicting Hot Leads

PROBLEM STATEMENT

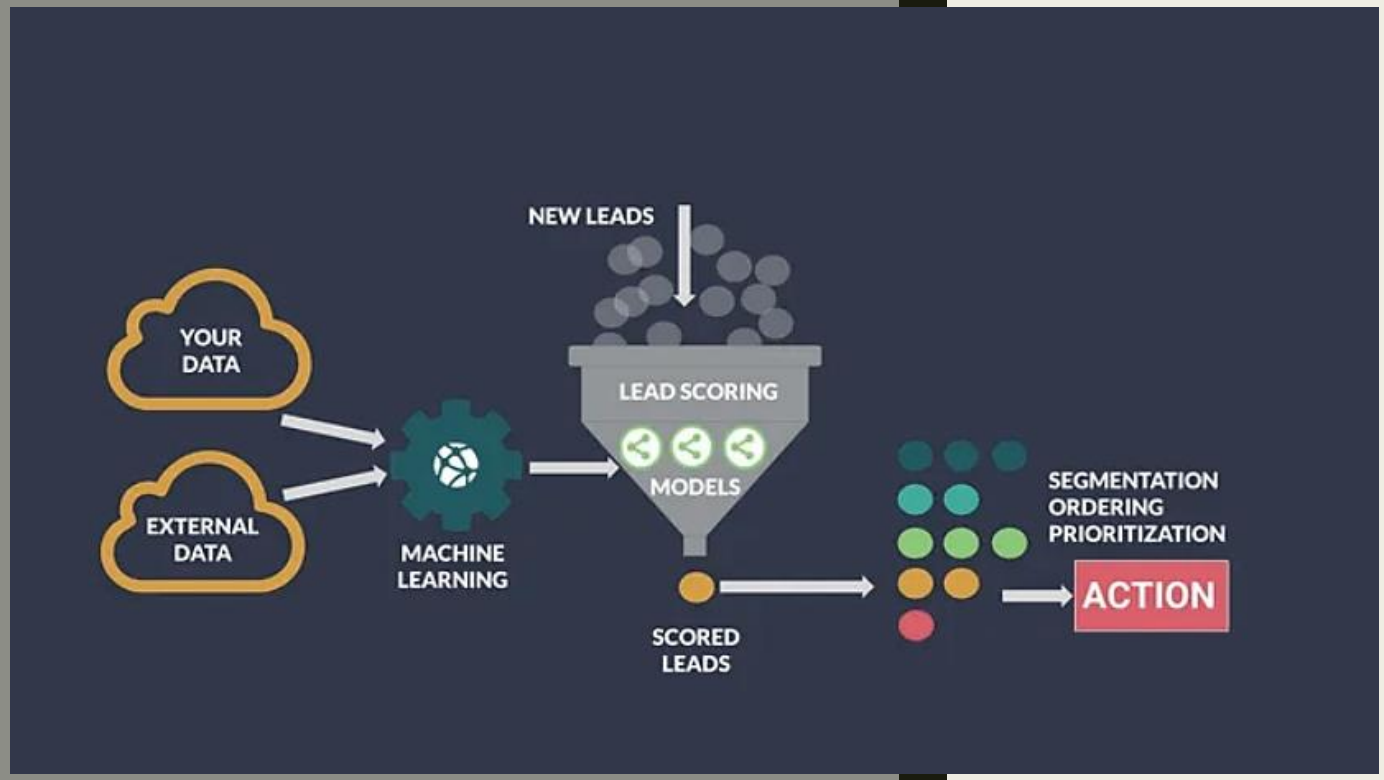
Businesses encounter a significant influx of leads from diverse channels, including website forms, referrals, and marketing campaigns.

To enhance conversion efficiency, businesses strive to identify the most promising leads, commonly known as 'Hot Leads,' who exhibit a greater likelihood of conversion.

This entails implementing strategies to identify, prioritize, and nurture leads with the highest potential for conversion. Businesses can significantly improve their overall lead conversion rate by focusing resources and efforts on these Hot Leads, increasing sales and revenue growth.



LEAD MANAGEMENT PROCESS



1 Lead Generation

2 Lead Qualification

3 Lead Conversion

WHAT HAS BEEN DONE?

- Existing methods in the current market, **Logistic regression** methods predict lead conversion and identify potential customers. These methods include traditional statistical approaches, rule-based systems, and machine-learning techniques.

WHAT ARE WE DOING?

- We are taking a data-driven approach to lead conversion prediction by implementing more advanced machine learning techniques, **regression,** and **clustering** along with **logistic regression**; we leverage these models to uncover complex patterns and relationships within the dataset, providing more accurate lead scoring prediction and providing **actionable insights** for **business decision-making**

GOAL

To develop a machine learning model that assigns a lead score to each potential Customer.

The objective is to optimize the lead conversion mechanism based on different parameters by Implementing explicit & implicit lead score modeling with a lead point system ranging from 0 to 100.

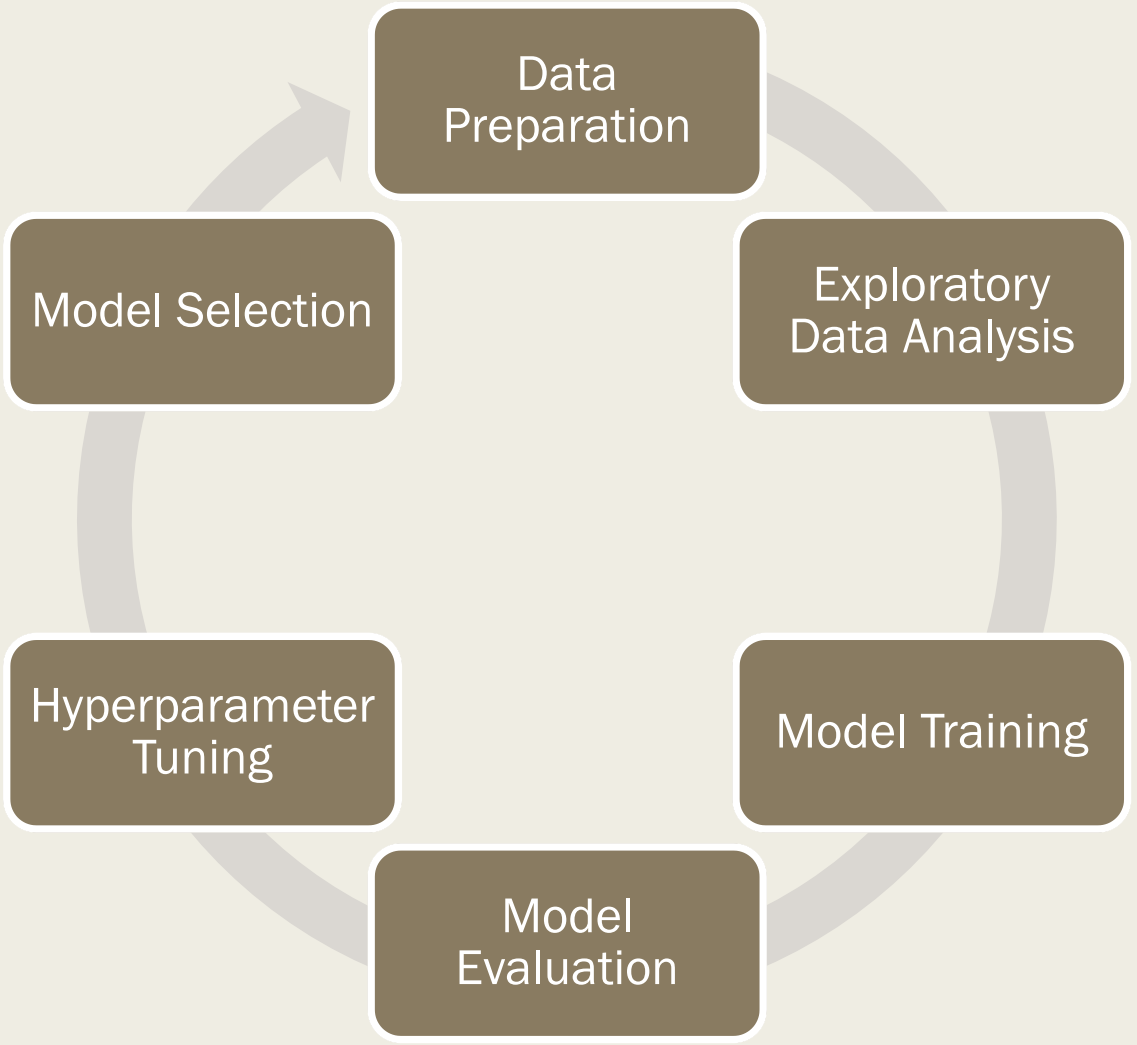
Variables	Description	Variables	Description
Prospect ID	A unique ID with which the customer is identified.	Specialization	The industry domain in which the customer worked before. Includes the level 'Select Specialization' which means the customer had not selected this option while filling the form.
Lead Number	A lead number assigned to each lead procured.	How did you hear about X Education	The source from which the customer heard about X Education.
Lead Origin	The origin identifier with which the customer was identified to be a lead. Includes API, Landing Page Submission, etc.	What is your current occupation	Indicates whether the customer is a student, umemployed or employed.
Lead Source	The source of the lead. Includes Google, Organic Search, Olark Chat, etc.	What matters most to you in choosing this course	An option selected by the customer indicating what is their main motto behind doing this course.
Do Not Email	An indicator variable selected by the customer wherein they select whether of not they want to be emailed about the course or not.	Search	Indicating whether the customer had seen the ad in any of the listed items.
Do Not Call	An indicator variable selected by the customer wherein they select whether of not they want to be called about the course or not.	Magazine	
Converted	The target variable. Indicates whether a lead has been successfully converted or not.	Newspaper Article	
TotalVisits	The total number of visits made by the customer on the website.	X Education Forums	
Total Time Spent on Website	The total time spent by the customer on the website.	Newspaper	
Page Views Per Visit	Average number of pages on the website viewed during the visits.	Digital Advertisement	
Last Activity	Last activity performed by the customer. Includes Email Opened, Olark Chat Conversation, etc.	Through Recommendations	Indicates whether the customer came in through recommendations.
Country	The country of the customer.	Receive More Updates About Our Courses	Indicates whether the customer chose to receive more updates about the courses.
I agree to pay the amount through cheque	Indicates whether the customer has agreed to pay the amount through cheque or not.	Tags	Tags assigned to customers indicating the current status of the lead.
a free copy of Mastering The Interview	Indicates whether the customer wants a free copy of 'Mastering the Interview' or not.	Lead Quality	Indicates the quality of lead based on the data and intuition the the employee who has been assigned to the lead.
Last Notable Activity	The last notable acitivity performed by the student.	Update me on Supply Chain Content	Indicates whether the customer wants updates on the Supply Chain Content.
Lead Profile	A lead level assigned to each customer based on their profile.	Get updates on DM Content	Indicates whether the customer wants updates on the DM Content.
City	The city of the customer.	Lead Profile	A lead level assigned to each customer based on their profile.

DATASET

We will be using a Lead Conversion dataset from Kaggle.

The dataset contains over **9,000 leads** with **31 customer features (variables)** such as lead origin, lead source, total time spent on the website, total visits on the website, demographics information, and the target column **Converted** (*indicating 1 for conversion and 0 for no conversion*).

APPROACH


Data Preparation
Exploratory Data Analysis
Model Training
Model Evaluation
Hyperparameter Tuning
Model Selection

IMPLEMENTATION PROCESS

- Preprocessing Data
- Data Split & Cross-Validation

Stage 1: Data Preparation & Analysis

(Ongoing Stage)

- Regression Models
- Classification Models
- Clustering Models

Stage 2: Model Building

- Evaluation Metrics
- Parameter Tuning
- Comparison & Selections

Stage 3: Model Selection & Tuning

STEP 1: DATA PREPROCESSING

Clean the data by handling missing values, outliers, and inconsistencies.

Use appropriate techniques for imputing missing values, such as mean or median imputation.

Detect and handle outliers using statistical methods like z-score or domain knowledge.

Apply feature scaling or normalization to ensure consistent ranges for different variables.

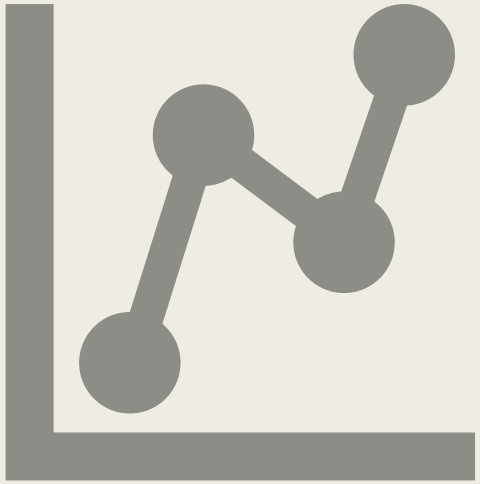
Explore and transform variables to meet regression, classification, or clustering algorithm assumptions.

STEP 2: DATA SPLIT & CROSS-VALIDATION

Split the preprocessed dataset into training (70%) and testing (30%) sets

Ensure the split maintains the distribution of target variables (stratified splitting)

Implement cross-validation techniques like k-fold or stratified cross-validation



STEP 3: REGRESSION MODELS

- Implement linear regression for modeling relationships between variables.
- Explore non-linear regression techniques like polynomial regression or decision tree regression.
- Apply regularization techniques like L1 (Lasso) or L2 (Ridge) regularization to control model complexity and prevent overfitting.
- Evaluate regression models using metrics like mean squared error (MSE), mean absolute percentage error (MAPE), or R-squared.

STEP 5: CLUSTERING MODELS

Apply

- Apply K-means clustering algorithm to identify potential customer segments based on lead characteristics.

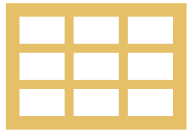
Implement

- Implement expectation maximization (EM) algorithm for clustering with probabilistic models.

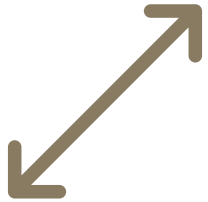
Utilize

- Utilize DBScan for density-based clustering and handling noise in the data.

STEP 4: CLASSIFICATION MODELS



Build logistic regression models for lead scoring and probability estimation.



Use regularization techniques like L1 or L2 to improve model performance and handle feature selection.

Explore Bayesian and Naive Bayes classification algorithms to capture probabilistic relationships.

Utilize support vector machines (SVM) with linear and non-linear kernels for classification tasks.

STEP 6: EVALUATION METRICS

Calculate accuracy, precision, recall, F1-score, and total accuracy for classification models.

Assess regression models using metrics like MSE, MAPE, or other appropriate regression evaluation metrics.

Evaluate clustering models using metrics like silhouette score or within-cluster sum of squares (WCSS).

STEP 7: PARAMETER TUNING

- Perform grid or random search for hyperparameter tuning in regression and classification algorithms.
- Use techniques like gradient descent (GD) or closed-form solutions for optimization in linear or non-linear regression models.

STEP 8: COMPARISON & SELECTION

Compare the performance of different models based on evaluation metrics.

Select the best-performing model for lead scoring, probability estimation, or customer segmentation based on the evaluation results.

INSIGHTS & CONCLUSION

The Machine Learning Model provides valuable insights and actionable recommendations for businesses aiming to predict and identify hot leads.

By leveraging these models, businesses can make data-driven decisions, enhance lead conversion rates, and drive successful marketing and sales efforts.

- Improve Lead Conversion
- Enhance Customer Segmentation
- Optimize Resource Allocation
- Drive Data-Driven Decision Making