

# LEAD SCORE FOR X EDUCATION USING LOGISTIC REGRESSION

---

BY DINESH BANAHATTI

# BUSINESS PROBLEM

---

- X Education is facing challenges with low conversion rates of leads into customers.
- Goal: Identify key factors influencing lead conversion and build a predictive model to improve targeting and conversion rates.

# DATASET SUMMARY

---

- The Leads.csv dataset contains:
  - Rows: 9,240
  - Features: 37
- Business Impact: The data can help identify which leads have the highest probability of converting into customers.
- Calculated the null values for each column.

# DATA PREPROCESSING

---

- Columns with over 30% missing data removed and also the columns that are not useful(e.g., City, Country).
- Irrelevant columns like 'Magazine,' 'Newspaper,' and 'Do Not Call' were dropped.
- Business Insight: Removing irrelevant data reduces noise, making the model more accurate in predicting conversions.

# FEATURE SELECTION AND ENCODING

---

- Dummy variables created for categorical columns (Lead Source, Specialization, etc).
- The selected features provide insights into what factors significantly impact the conversion.

# SCALING SPLITTING DATA

---

- Imported MinMax.
- The column like 'TotalVisits','Total Time Spent on Website','Page Views Per Visit' applied scaling.
- Min-Max Scaling standardizes data for machine learning models, improving accuracy and efficiency in business decision-making by ensuring all features contribute equally to predictions.
- Split the dataset into test and train in 70% and 30% respectively.



# LOGISTIC REGRESSION APPROACH

---

- Logistic Regression is used to model the probability of lead conversion (1 = Converted, 0 = Not Converted).
- Used Recursive Feature Elimination (RFE) to select top 15 most relevant features.
- Recursive Feature Elimination helped us identify the top predictive features.
- Business Insight: This approach helps pinpoint the exact features that influence conversion, allowing for better resource allocation in lead nurturing.

# VIF AND PVALUE

---

- Use statsmodels to fit a regression model.
- Calculate VIF and check p-values via `model.summary()`.
- Helps ensure robust models for effective business decision-making.



# MODEL PERFORMANCE ON TRAINING DATA

---

- Accuracy: 71%
- Sensitivity (Recall): 65%
- Specificity: 77%
- ROC Curve: AUC score of 0.79
- Business Insight: The model performs well on training data, meaning it can effectively differentiate between converting and non-converting leads.

# MODEL PERFORMANCE ON TEST DATA

---

- Accuracy: 71%
- Sensitivity: 65%
- Specificity: 75%
- Business Insight: The model generalizes well to unseen data, indicating it can be reliably used in production for lead scoring.

# STRATEGIC RECOMMENDATIONS

---

- **Optimize Lead Generation Channels:** Enhance the **Lead Add Form** and improve visibility on the **Welingak Website** to attract high-quality leads.
- **Tailor Communication Strategies:** Develop SMS and phone call campaigns for leads marked as **Do Not Email** to increase engagement.
- **Segment Marketing Efforts:** Create targeted marketing campaigns addressing the specific needs of **Unemployed** and **Student** leads to improve conversion rates.
- **Monitor and Adjust Based on Engagement Metrics:** Implement tracking systems for engagement metrics to adapt strategies based on lead behavior.
- **Regularly Review and Update the Model:** Continuously reassess and refine the logistic regression model to ensure relevance and effectiveness in changing market conditions.

# FINAL THOUGHTS

---

- Predictive model helps X Education identify high-potential leads, improving sales efficiency.
- The model provides actionable insights for targeted marketing and resource allocation.
- Next steps: Continuously update the model with new data and track performance in real-time.