NEURONETIX

EXPLORE THE FUTURE

ABDELGHAFOR'S HACKATHON

MACHINE LEARNING HACKATHON

Telecom Customer Churn Prediction and Segmentation

Project Overview:

This project focuses on analyzing customer behavior in the telecom industry to predict churn and identify meaningful customer segments. Using classical machine learning techniques, the goal is to build predictive models that classify customers into churn or non-churn categories and segment them into groups based on shared characteristics. This analysis will empower telecom businesses to improve customer retention and tailor services to different customer segments.

Key Features:

1. **Churn Prediction**:

   - **Supervised Learning**: Develop models to classify customers as likely to churn or remain, using features like service usage, contract type, and payment methods.

   - **Feature Engineering**: Extract and create new features to enhance the accuracy of the predictive model.

   - **Model Comparison**: Apply multiple algorithms (Logistic Regression, Decision Trees, Random Forest, etc.) and select the best model based on evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

2. **Customer Segmentation**:

   - **Unsupervised Learning**: Implement clustering algorithms (K-Means, Hierarchical Clustering) to segment customers based on their behavior and demographics.

   - **Optimal Clustering**: Use techniques like the elbow method or silhouette score to determine the best number of clusters, ensuring meaningful segmentation.

   - **Insights**: Analyze customer segments and identify patterns such as high-risk churn segments or high-value customers.

3. **Visualization & Reporting**:

   - o Visualize the churn prediction results, customer clusters, and their characteristics using Python libraries such as Matplotlib and Seaborn.

   - o Present insights and recommendations in a structured format, emphasizing actionable business strategies.

Project Requirements:

1. **Python**: Use Python for data analysis, model building, and evaluation.

2. **Libraries**:

   - o **Pandas** and **NumPy** for data manipulation.

   - o **Scikit-learn** for machine learning algorithms.

   - o **Matplotlib** and **Seaborn** for data visualization.

3. **Machine Learning Techniques**:

   - o **Supervised Learning**: Apply at least two supervised learning algorithms for churn prediction.

   - o **Unsupervised Learning**: Apply at least one clustering algorithm for customer segmentation.

4. **Data Preprocessing**: Handle missing data, encode categorical variables, scale numerical variables, and ensure the dataset is prepared for machine learning.

5. **Evaluation Metrics**: Include metrics such as accuracy, precision, recall, F1-score, and ROC-AUC for classification. For clustering, use silhouette score and other relevant metrics to evaluate clustering quality.

6. **Presentation**: Prepare a comprehensive presentation summarizing your findings, models, and insights.

Criteria for Evaluation:

1. **Data Preprocessing (20%)**:

    o   Handling missing data.

    o   Proper encoding of categorical features.

    o   Scaling of numerical features.

    o   Feature engineering and selection.

2. **Supervised Learning Model (30%)**:

    o   Accuracy and performance of the churn prediction model.

    o   Proper usage and comparison of multiple algorithms.

    o   Model evaluation and justification of the selected model.

3. **Unsupervised Learning Model (20%)**:

    o   Implementation of clustering algorithms.

    o   Selection of the optimal number of clusters.

    o   Insights derived from customer segments.

4. **Data Visualization (15%)**:

    o   Clear and informative visualizations for both the prediction model and clustering results.

    o   Effective use of plots to support insights and recommendations.

5. **Presentation (5%)**:

    o   Structured and clear presentation of project findings.

    o   Explanation of the models, processes, and business implications.

    o   Engagement and communication of results.

6. **Overall Analysis & Insights (10%)**:

    o   Quality and depth of business insights derived from the analysis.

    o   Actionable recommendations for reducing churn and improving customer retention.

Bonus Points (Optional):

1. **Advanced Feature Engineering**:

   o Using domain-specific knowledge to create impactful features.

   o Application of techniques like PCA (Principal Component Analysis) for dimensionality reduction.

2. **Exploring Different Clustering Algorithms**:

   o Beyond K-Means, using advanced clustering techniques like DBSCAN or Hierarchical Clustering.

3. **Advanced Visualization**:

   o Interactive visualizations (e.g., using Plotly) for customer segment insights.

4. **Novel Insights**:

   o Providing unique and actionable insights that go beyond standard churn prediction, such as identifying upselling opportunities or suggesting specific retention strategies for each customer segment.

NEURONETIX

EXPLORE THE FUTURE

GOOD LUCK