BEHAVIOR-DRIVEN CUSTOMER SEGMENTATION AND CHURN PREDICTION USING HYBRID CLUSTERING AND EXPLAINABLE AI

R.AVINASH REDDY

22P61A05K1

reddyavinash498@gmail.com

TANISHQ VERMA

22P61A05N3

tanishqverma193@gmail.com

T.KEERTHANA

22P61A05N4

keerthanathabety@gmail.com

ABSTRACT :

This project proposes a hybrid, intelligent Customer Segmentation and Churn Prediction System that enhances business decision-making through machine learning and explainable AI. By combining RFM modeling with advanced clustering techniques such as K-Means, Gaussian Mixture Models, and DBSCAN, customers are segmented based on their behavioral patterns. Additionally, a supervised learning module predicts customer churn using classification models. The system integrates SHAP-based Explainable AI to offer transparency behind predictions and clustering decisions. A dynamic dashboard provides visual insights, churn probabilities, and actionable business strategies. The solution presents a robust and scalable approach to customer analytics, blending unsupervised and supervised learning with interpretability and user-friendly insights.

INTRODUCTION :

Understanding customer behavior is crucial for businesses aiming to enhance customer satisfaction and improve marketing strategies. Traditional analysis methods often fall short in identifying nuanced storage services face challenges related to consumer patterns. This project introduces an advanced system that leverages unsupervised learning (RFM + hybrid clustering) and supervised models (churn prediction), combined with Explainable AI (XAI) using SHAP. By revealing actionable insights, customer segmentation becomes more precise and interpretable, allowing for highly targeted marketing and retention efforts.

HARDWARE & SOFTWARE REQUIREMENTS :

• Processor: Multi-core processor (e.g., Intel Core i5/i7 or equivalent)
• RAM: Minimum 8 GB for smooth data processing
• Storage: Minimum 250 GB free disk space
• Network: Stable internet connection for dataset access and model deployment.
• Operating System: Windows / Linux / macOS
• Programming Language: Python
• Libraries/Tools: NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, XGBoost, SHAP, Streamlit
• IDE: Jupyter Notebook / VS Code
• Database (optional): SQLite or CSV files for storing and retrieving customer data

EXISTING SYSTEM :

Conventional customer analysis often relies on manual categorization and rule-based systems that lack flexibility and scalability. These systems typically fail to identify subtle variations in customer behavior, resulting in generalized marketing approaches and reduced effectiveness. They also don't incorporate churn analysis or interpretable ML, leading to under-utilized data and missed business opportunities.

REFERENCES :

1. Tan, S., & Kumar, A. (2020). "Clustering Techniques for Customer Segmentation: A Comparative Study." International Journal of Data Science, 8(2), 155–162.

2. Smith, R., & Zhao, Y. (2021). "Enhancing Retail Marketing with RFM Analysis and Machine Learning." Retail AI Conference Proceedings.

3. Li, J., & Chen, M. (2019). "Unsupervised Learning in E-commerce: Applications in Customer Behavior Analysis." Journal of Machine Learning Applications, 5(1), 72–80.

4. Lundberg, S. M., & Lee, S.-I. (2017). "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems.

GUIDE NAME : Dr. N. Srinivas
PROJECT COORDINATOR:
HOD: