IMPROVING PREDICTION ACCURACY WITH ADVANCE MACHINE LEARNING TECHNIQUE

TEAM MEMBERS:

1. SENTHIL VEL S (LEADER)
2. VIGNESH M
3. GUNASEKARAN P
4. GUNASEKARAN S
5. KAMALESH K

IMPROVE CUSTOMER CHURN PREDICTION ACCURACY:

- Incorporating advanced machine learning techniques like ensemble models .
- feature engineering can indeed improve the prediction accuracy of customer churn models.
- Customer churn prediction is a critical task for businesses, as it helps in retaining customers and increasing profitability.

ENSEMBLE MODELS:

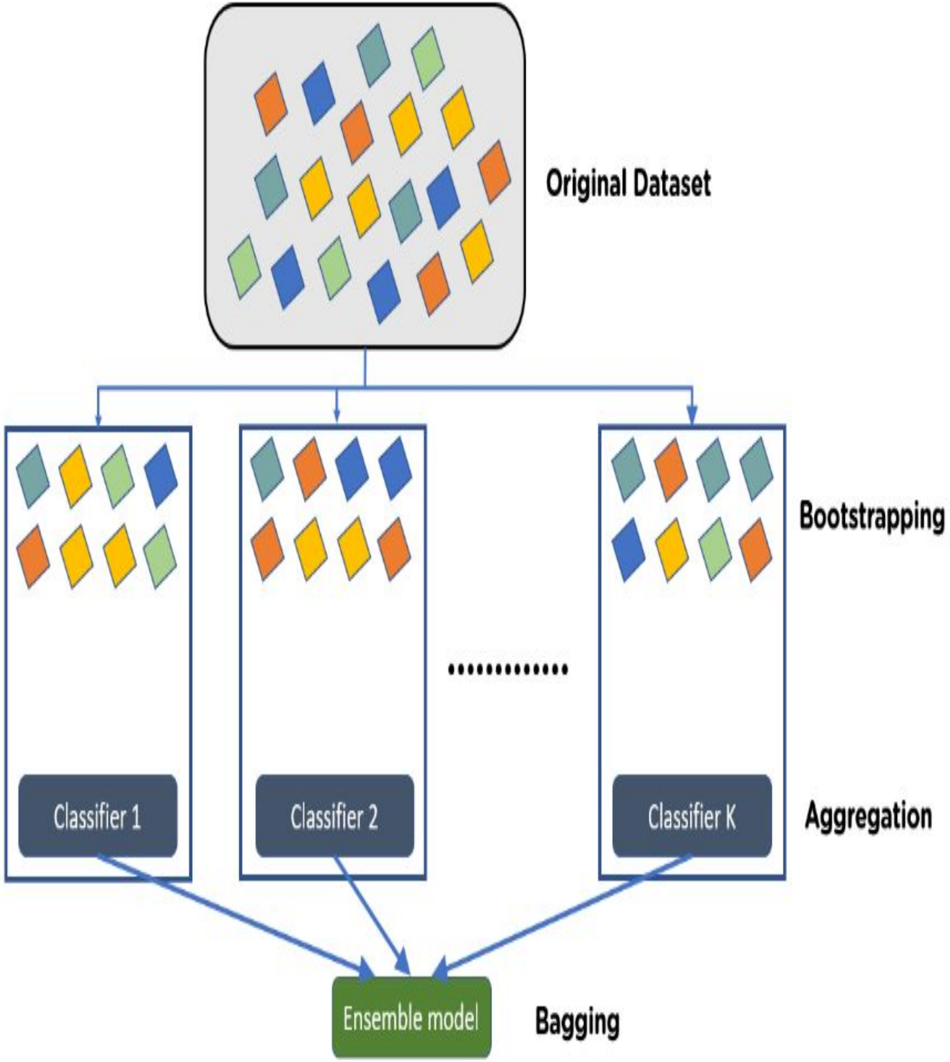
- Ensemble modeling is a process where multiple diverse models are created to predict an outcome, either by using many different modeling algorithms or using different training data sets.
- The ensemble model then aggregates the prediction of each base model and results in once final prediction for the unseen data.

Ensemble model techniques :

- **Bagging (Bootstrap Aggregating):** Use algorithms like Random Forest to train multiple decision trees on different subsets of the data and aggregate their predictions.
- **Boosting:** Algorithms like AdaBoost, Gradient Boosting, or XGBoost can be used to sequentially train models, giving more weight to misclassified instances, which helps in improving accuracy.
- **Stacking:** Combine predictions from multiple diverse models (e.g., decision trees, SVMs, neural networks) using another model (meta-learner) to make the final prediction.

STEP TO PERFORM BAGGING :

- Consider there are n observations and m features in the training set. You need to select a random sample from the training dataset without replacement
- A subset of m features is chosen randomly to create a model using sample observations
- The feature offering the best split out of the lot is used to split the nodes
- The tree is grown, so you have the best root nodes
- The above steps are repeated n times. It aggregates the output of individual decision trees to give the best prediction



FEATURE ENGINEERING :

- Feature engineering is a machine learning technique that leverages data to create new variables that aren't in the training set.

- It can produce new features for both supervised and unsupervised learning, with the goal of **simplifying and speeding up data transformations** while also **enhancing model accuracy**

FEATURE:

Demographic Information:

- Age: The age of the customer can be a relevant factor. Younger customers might be more tech-savvy or less risk-averse.
- Gender: Gender can also play a role in predicting churn, as preferences may vary by gender.
- Income: Higher-income customers might be less price-sensitive and less likely to churn.

Usage Patterns:

- Frequency of Use: How often a customer uses the service or product can be a strong indicator. Frequent users are less likely to churn.
- Recency of Use: When was the last time the customer used the service? Inactive users may be more likely to churn.
- Usage Volume: The amount or volume of product or service used can be a valuable feature. Higher usage can indicate customer satisfaction.

FEATURE:

Billing and Payment Information:

- Payment History: Analyzing payment histories, such as late payments or bounced payments, can be indicative of a churn risk.
- Billing Plan: The type of billing plan a customer is on can impact their likelihood of churning. For example, customers on long-term contracts might be less likely to churn.

Customer Interactions:

- Customer Support Interactions: The number of interactions with customer support can be an important signal. Frequent support requests might indicate dissatisfaction.
- Complaints or Feedback: Customer feedback, complaints, or survey responses can provide valuable insights into customer satisfaction.

WORKING....

Data Collection:

- Gather a comprehensive dataset that includes relevant information about customers, their interactions with your business, and churn outcomes.
- Gather relevant data: Collect historical customer data that includes customer attributes, transaction history, usage patterns, and churn labels.

Data Exploration:

- Begin by examining the dataset to understand its structure and characteristics.
- Check for missing data, outliers, and data imbalances.

Data Preprocessing:

- Clean and preprocess the data. This includes handling missing values, removing duplicates, and addressing outliers.
- Convert categorical variables into numerical format (e.g., one-hot encoding or label encoding).
- Scale or normalize numerical features, so they have similar scales and do not dominate the modeling process.

WORKING....

Feature Engineering:

- Create new features or transform existing ones to extract valuable insights. Some common features for churn prediction include customer lifetime value, churn risk scores, and engagement metrics.

Data Splitting:

- Divide the dataset into training, validation, and test sets. The training set is used to train the model, the validation set is used for hyperparameter tuning and model selection, and the test set is used to evaluate the model's performance.

Model Selection:

- Choose appropriate machine learning algorithms and models. Common choices for customer churn prediction include logistic regression, decision trees, random forests, support vector machines, and neural networks.

WORKING....

Monitoring and Updating:

- Continuously monitor the model's performance and update it as needed. Customer behavior and churn patterns can change over time.

Feedback Loop:

- Create a feedback loop to learn from the outcomes of your retention efforts and improve the model accordingly.

Cross-Validation:

- Use cross-validation techniques to assess the model's stability and generalization performance. K-fold cross-validation is a common choice.

WORKING....

Feature Selection:

- Identify and select the most relevant features. Feature selection helps avoid overfitting and improves model interpretability.

Regularization:

- Apply regularization techniques (e.g., L1 or L2 regularization) to prevent overfitting, especially for complex models.

Ensemble Models:

- Consider using ensemble methods, such as stacking or bagging, to combine predictions from multiple models for improved accuracy and robustness.

Model Interpretability:

- Ensure that your model is interpretable and can provide explanations for its predictions. This is valuable for understanding why customers are likely to churn.

WORKING....

Model Training:

- Train the selected model using the training dataset. Experiment with different hyperparameters to optimize the model's performance.

Model Evaluation:

- Assess the model's performance using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, ROC AUC, or others, depending on the problem's specific requirements.
- Conduct cross-validation to ensure the model's generalization ability.

Hyperparameter Tuning:

- Fine-tune the hyperparameters of the model to improve its performance. This can be done using techniques like grid search, random search, or Bayesian optimization.

WORKING....

Testing on Unseen Data:

- After model selection and training, evaluate its performance on the test set to get an accurate estimate of its real-world predictive power.

Testing:

- Implement tests to validate the effectiveness of your predictions and interventions. This can help determine the real impact of your churn prediction efforts.

THANK YOU !