***PHASE 1***

***PREDICTING CUSTOMER CHURN USING MACHINE LEARNING TO UNCOVER HIDDEN PATERNS***

***STUDENT NAME: SATHISH KUMAR .S***

***REGISTER NUMBER: 511723104010***

***INSTITUTION: PALLAVAN COLLEGE OF ENGINEERING***

***DEPARTMENT: COMPUTER SCIENCE AND ENGINEERING***

***DATE OF SUBMISSION: 01/05/2025***

# ***1.Problem Statement: Predicting Customer Churn Using Machine Learning***

*Stock prices fluctuate based on various economic, political, and company-specific factors. Predicting future stock movements is a complex task that requires analyzing time series data and understanding market trends. The goal of this project is to build a machine learning model that uses historical stock data to forecast future stock prices. This will assist investors in making data-driven financial decisions and improve their risk assessment strategies*.

***2. Data Collection***

*Use historical stock price data (Open, High, Low, Close, Volume).*

*Source data using:*

*yfinance Python library*

*Alpha Vantage or IEX Cloud APIs*

*Kaggle datasets*

***3. Data Preprocessing***

*Data Cleaning: Handle missing prices, remove duplicates, align timestamps.*

*Feature Engineering:*

*Generate lag features (e.g., yesterday’s closing price).*

*Create technical indicators (SMA, EMA, RSI, MACD).*

*Scaling: Apply MinMaxScaler or StandardScaler on numeric features.*

*Stationarity: Apply transformations (differencing/log) if needed for time series models.*

***4. Exploratory Data Analysis (EDA)***

*Plot price trends over time.*

*Analyze daily, weekly, and monthly volatility.*

*Compute and visualize correlation matrices between technical indicators and closing price.*

*Identify seasonal patterns and long-term trends.*

***5. Model Selection***

* *Choose and compare multiple models:*
* *Statistical Models: ARIMA, SARIMA*
* *Machine Learning Models: Linear Regression, Random Forest, XGBoost*
* *Deep Learning Models: LSTM, GRU (for sequence learning)*

***6. Model Training***

*Split data into training and testing sets based on time.*

*Use time series cross-validation (e.g., expanding window).*

*Tune hyperparameters using GridSearchCV or Bayesian optimization (where applicable).*

*Handle model performance degradation by retraining on new data.*

***7. Model Evaluation***

*Split data into training and testing sets based on time.*

*Use time series cross-validation (e.g., expanding window).*

*Tune hyperparameters using GridSearchCV or Bayesian optimization (where applicable).*

*Handle model performance degradation by retraining on new data.*

***8. Feature Importance & Explainability***

* *Use SHAP or Permutation Importance to evaluate which technical indicators contribute most to price predictions.*

***9. Model Saving***

*Save the best-performing model for future use:*

*python*

*Copy*

*Edit*

*import joblib*

*joblib.dump(model, 'stock\_price\_predictor.pkl')*

***10. Model Deployment***

*Option 1: Flask/FastAPI*

*Create a REST API that receives dates or recent stock data and returns predicted prices.*

*Option 2: Streamlit*

*Build an interactive dashboard for visualizing predictions and uploading stock symbols.*

*Option 3: Cloud Deployment*

*Host on platforms like AWS EC2, Heroku, or GCP App Engine*.

***11. Monitoring & Maintenance***

*Track model performance over time.*

*Set up scheduled retraining using new stock data.*

*Implement alerting if model drift is detected.*

***Team members , roles and responsibilities***

***1. Project Manager [ KEERTHANA ]***

* ***Responsibilities:***
  + *Define project scope, goals, and timeline.*
  + *Coordinate tasks among team members.*
  + *Ensure communication with stakeholders.*
  + *Track project progress and ensure deadlines are met.*
  + *Assist with high-level decision-making and risk management.*

***2. Data Engineer [ NAZREEN FARZANA]***

* ***Responsibilities:***
  + *Collect and integrate data from various sources (databases, APIs, files).*
  + *Clean, preprocess, and transform raw data into usable formats.*
  + *Handle missing values, outliers, and data normalization.*
  + *Design data pipelines for future automation and scalability.*

***3. Data Scientist [ ANITHA ]***

* ***Responsibilities:***
  + *Perform exploratory data analysis (EDA) to identify trends and patterns.*
  + *Select and engineer features for model input.*
  + *Build, train, and evaluate machine learning models.*
  + *Tune hyperparameters and optimize model performance.*
  + *Conduct feature importance and model explainability (e.g., SHAP, LIME).*

***4. Backend Developer [ DIVYA ]***

* ***Responsibilities:***
  + *Develop APIs (e.g., using Flask or FastAPI) to serve the ML model.*
  + *Handle model versioning and deployment (e.g., using Docker, Kubernetes).*
  + *Integrate the model into a production-ready system.*
  + *Monitor model performance and set up retraining pipelines.*
  + *Ensure security and scalability of the deployed system.*

***5. Dashboard designer [ ANANTHI ]***

* ***Responsibilities:***
  + *Design and build an intuitive front-end interface (e.g., using Streamlit, React).*
  + *Create interactive visualizations of churn predictions and model insights.*
  + *Enable business users to upload data and receive predictions.*
  + *Ensure a user-friendly, responsive, and accessible design.*