"PRACTICAL NO. 06"

AIM :

Predict the credit worthiness of customer / credit card fraud detection.

INPUTS :

- Transaction data,
- customer profile / customer data
- Previous labels [if available]

EXPECTED OUTPUTS :

- credit worthiness
- Fraud detection

THEORY :

The theory behind predicting credit worthiness or detecting fraud involves using classification algorithms that can analyze patterns within the data. Fraud detection models, in particular, need to be highly accurate due to the costs associated with false negatives i.e. fraudlent transactions incorrectly classified as legitimate.

creditworthiness prediction is typically a credit scoring problem, where a model evaluates a customer's likelihood to repay a loan or credit. The scoring may consider a variety of factors including credit history, income and spending behaviour.

Fraud detection is a binary classification problem that involves identifying anomalies in transaction data. The challenge here lies in handling imbalance data since fraudul-ent transactions are generally much rarer than legitimate ones.

Additionally, feature engineering plays a vital role in both prediction task. Creditworthiness models benefit from features like credit utilization ratio and recent changes in the spending behaviour. For fraud detection, useful features might include the number of transactions location and transaction time-of-day patterns. Creating meaningful features helps model identify the most relevant patterns in customer or transaction data, ultimately improving prediction accuracy. Proper feature scaling and encoding are essential to ensure are essential to ensure that these models learn correctly and efficientley from the data.

In both creditworthiness and fraud detection, model interpretability is an important consideration. Financial institutions often prefer interpretable models like decision trees or logistic regression because these provide clear explanation for why a certain decision was made. This is particularly critical for regulatory compliance and customer trust.

For instance, a customer should be able to understand why their credit application was denied, while transparency in fraud detection decisions can help business refine their strategies to prevent fraudulent behaviour.

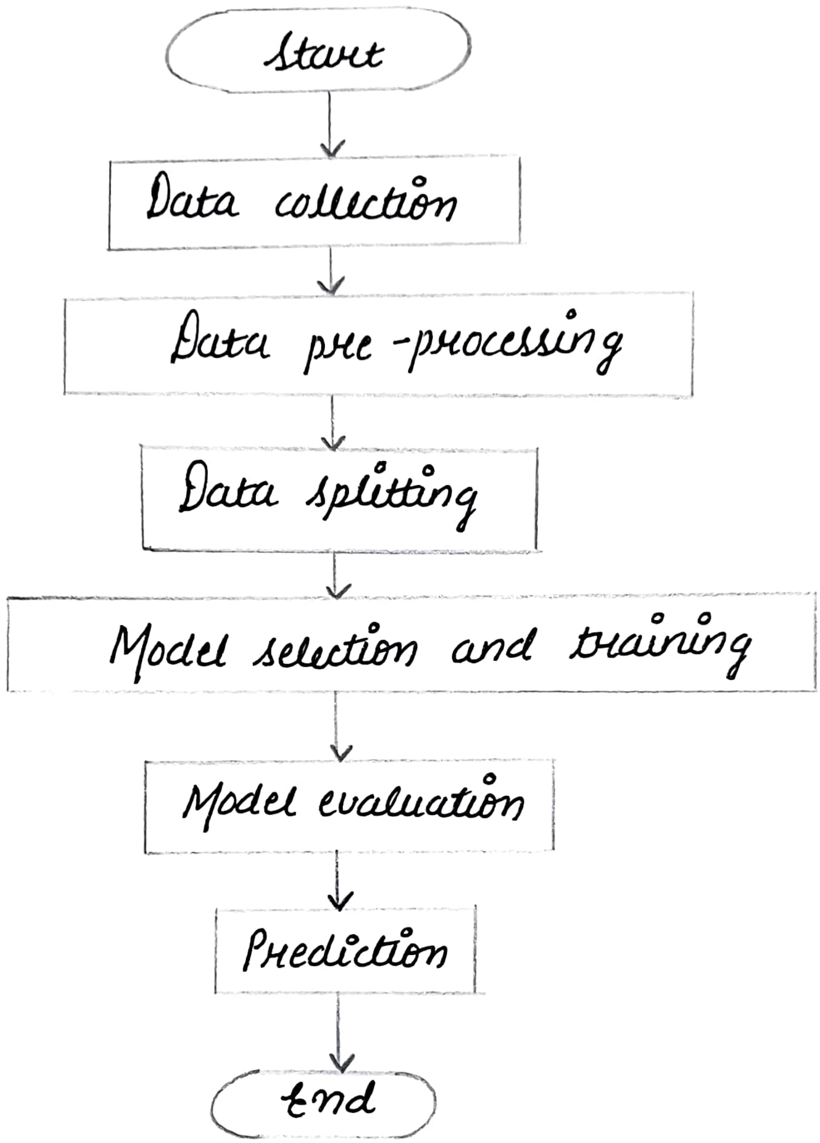
ALGORITHM :

STEP [1] : Gather data on customer or transactions.
STEP [2] : Handle missing values, normalize numerical data, encode categorical values.
STEP [3] : Divide the dataset into training set and test set.
STEP [4] : Train a random forest classifier on the training data.
STEP [5] : Use metrics such as accuracy, precision, Recall and F1 - score.
STEP [6] : Apply the trained model on the new customer data.

CONCLUSION :
This demonstrates the use of machine learning for both creditworthiness prediction and fraud detection. The results help in identifying customers with a higher risk of default or flagging suspicious transactions. Machine learning model like Random forest are well - suited for these tasks due to their accuracy and interpretability.

[P-T-O]

Through effective feature selection, data preprocessing and model tuning, the implemented algorithm can reduce credit risk and prevent fraudulent transactions, benefiting both financial institutions and customers by enhancing security and reliability in credit systems.


Start
Data collection
Data pre-processing
Data splitting
Model selection and training
Model evaluation
Prediction
End

"PRACTICAL NO. 07"

AIM :

Predict the sentiment from social media or customer review

INPUTS :

The input to the system consists of text data obtained from:
- Social media posts [e.g. Twitter, Facebook]
- Customer reviews [e.g. Amazon, Google reviews]

EXPECTED OUTPUT :

The output is a sentiment prediction, where each input text is classified into - positive, negative, neutral.

THEORY :

Sentiment analysis leverages NLP techniques to analyze the sentiment or emotional tone of text data. Key theoretical concepts include -

- Text Preprocessing :
  Involves cleaning and normalizing text [e.g. tokenization, stop-word removal, stemming and lemmatization] to improve analysis accuracy.
- Vectorization :
  Converting text into a numerical format that algorithms can understand.

Common techniques include :
i] Bag of words [BoW]
ii] Term frequency - inverse document Frequency [TF - IDF]
iii] word Embeddings [e.g. Word2Vec , GloVe]

- Classification algorithms : Machine learning algorithms used to predict sentiment, such as Naive Bayes, Support Vector Machines and neutral networks.
- Evaluation metrics : Metrics like accuracy, precision, recall and F1 score used to evaluate model performance.

Sentiment analysis, or opinion mining is a Natural Language Processing [NLP] technique used to determine the emotional tone within a body of text. This technique is often applied to online reviews, social media post or any other form of customer feedback to categorize opinions as positive, negative or neutral.

By analyzing text data for sentiment, companies and researchers can obtain valuable insights into customer attitudes, brand perception and emerging trends in consumer sentiment. This task is complex due to the nuances of human language where sarcasm, slang and idiomatic expressions can change the meaning of sentence presenting challenges for automated analysis.

A fundamental part of sentiment analysis is text processing, which involves cleaning and normalizing the raw text data to make it more suitable for analysis. Text pre-processing steps typically include removing punctuation, numbers and special charecters that do not contribute to sentiment. Techniques as tokenization like removing common words that carry little meaning like 'the' and 'is' and stemming or lemmatization are commonly used.

Through careful evaluation we can fine-tune the model, improving its generalizability and robustness for real-world applications.

ALGORITHM :

STEP [1] : Gather text data from social media or review sources.

STEP [2] : Remove unnecessary symbols, punctuations and also special charecters.

STEP [3] : Convert the processed text into numerical vectors

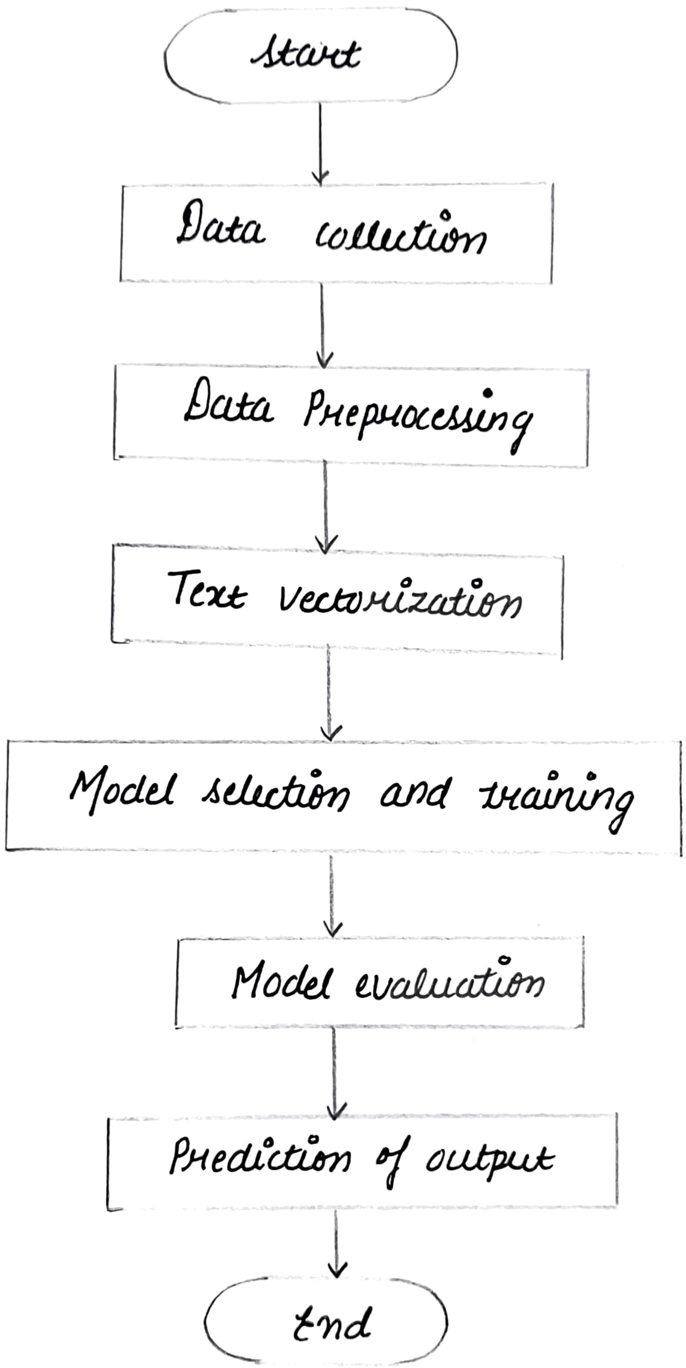
STEP [4] : Choose the machine learning model.

STEP [5] : Evaluate the model on test data using metrics.

STEP [6] : Use the trained model to predict the sentiment.

CONCLUSION :

sentiment analysis provides an efficient way to process large volumes of text data, allowing business and researchers to derive insights from user feedback and social media trends. By using machine learning and NLP techniques, the model can accurately classify sentiments, providing a valuable tool for improving customer satisfaction and understanding public opinion.


Start
Data collection
Data Preprocessing
Text vectorization
Model selection and training
Model evaluation
Prediction of output
End

"PRACTICAL NO. 08"

AIM :
Predict the future energy consumption for household or industry based on past data.

INPUTS :
- Historical Energy Consumption data
- Time variables
- Weather details

EXPECTED OUTPUT :
Time series prediction of future energy consumption.

THEORY :
Energy consumption prediction is based on the principle that historical data holds valuable insights into future patterns. Energy consumption patterns often reflect the influence of various factors like time of day, season, temperature and human activity. For instance, households and industries tend to consume more energy in colder months for heating and in warmer months for cooling. similarly, weekends and holidays may see a different usage pattern compared to week days. By analyzing trends, predictive models can learn from past data and projects future consumption.

Time series analysis is central to energy consumption forecasting. Time series models, such as ARIMA [Auto Regressive Integrated Moving Average] and SARIMA [Seasonal ARIMA] are widely used in such predictions due to their effectiveness in handling data with seasonality and trends. ARIMA models assume that future value depend linearly on past values, forecast errors, making them siutable for linear, stationary data. These models are effective for capturing simple and stable time dependencies but may struggle with non-linear or highly fluctuating data for model.

Machine learning models like linear regression and advanced neural networks such as LSTM [Long Short-term memory] have gained popularity in energy forecasting. Linear regression is useful for capturing linear relationships between energy consumption. However, when dealing with non-linear, complex pattern, neural networks like LSTM, a type of recurrent nural network [RNN] are particularly advantageous.

To evaluate model performance, common metrics like Mean Absolute Error [MAE], Mean Squared Error [MSE] and Root Mean Squared Error [RMSE] are used. These metrics quantify how closely the prediction align with actual values, helping to select the most suitable model for the data.

cross validation and hyperparameter tuning are also employed to optimize model performance. Once validated, the model can be deployed for real-time forecasting, continuously updated as new data arrives. Accurate predictions are highly beneficial in guiding efficient energy allocation, helping power companies adjust their production schedules are supporting households and industries in better managing energy consumption.

ALGORITHM :

STEP[1]: Gather historical data on energy consumption along with auxiliary data.

STEP[2]: Handle missing values by filling, interpolating or removing them.

STEP[3]: Visualize data trends, seasonality and anomalies and checking for correlations between energy consumption and other factors.

STEP[4]: Choose suitable model for forecasting.

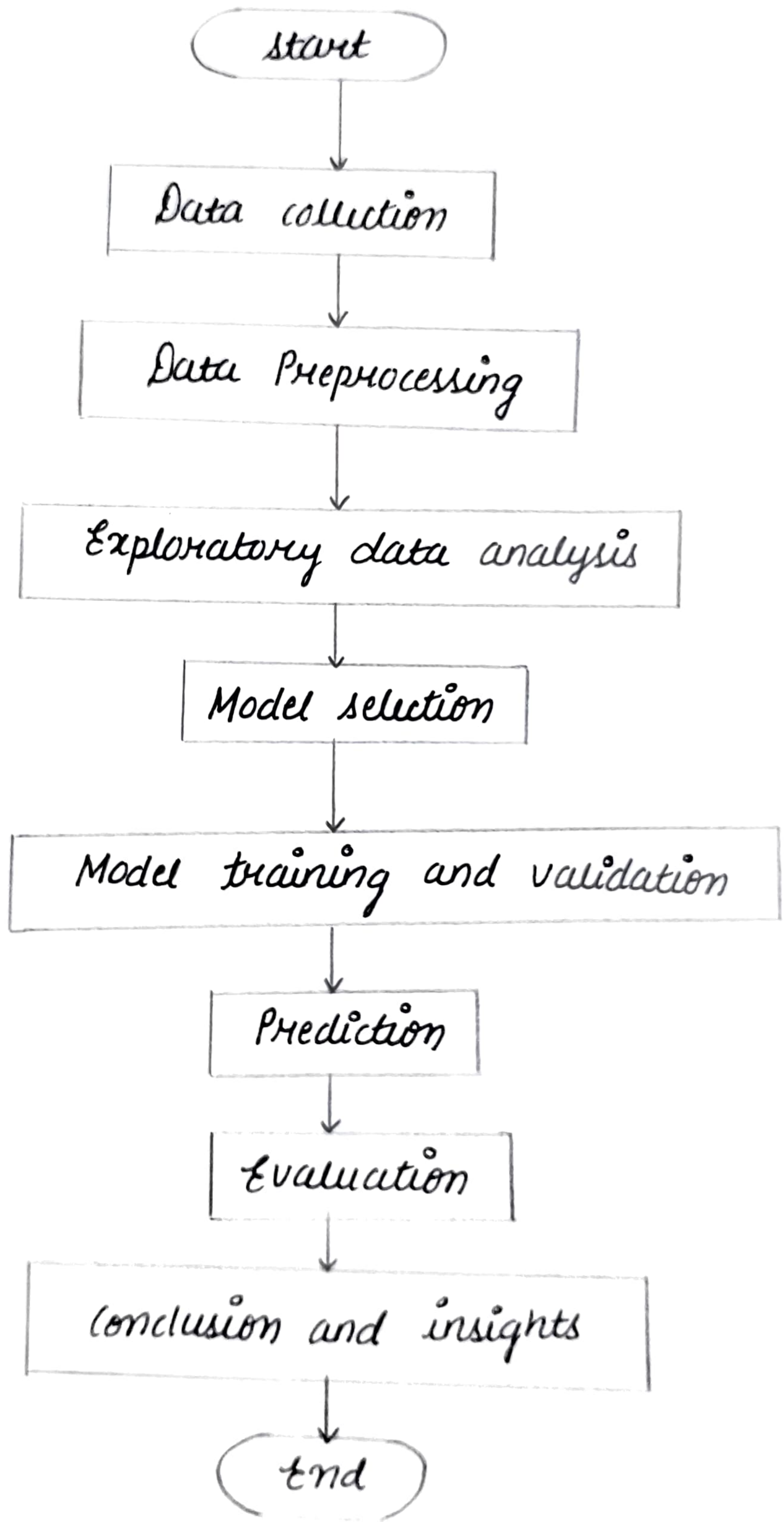
STEP[5]: Split data into test set and training set.

STEP[6]: Generate future predictions based on the chosen model.

STEP[7]: Retrain with new data or adjust parameters to improve prediction accuracy.

CONCLUSION :

This practical demonstrates how historical data can be leveraged to predict future energy consumption, which is essential for effective energy planning and management. By selecting an appropriate model and turning it to fit the patterns in the data, we can make accurate predictions. Real - world challenges such as missing data, seasonality and changing consumption patterns need to be addressed to ensure robustness. Accurate predictions allow for more efficient energy resource allocation, cost savings and support for sustainable energy strategies.


start
Data collection
Data Preprocessing
Exploratory data analysis
Model selection
Model training and validation
Prediction
Evaluation
Conclusion and insights
End