**BANK RISK CONTROLLER SYSTEMS**

**METHODOLOGY**

1. **Data Collection**:
   * Collect relevant banking data from various sources, such as transaction records, customer information, credit scores, and market data.
   * Ensure the data is clean, complete, and in a suitable format for analysis.
2. **Data Preprocessing**:
   * Perform data preprocessing techniques, such as data normalization, feature scaling, and encoding categorical variables.
   * Handle missing values and outliers using appropriate methods (e.g., mean imputation, median imputation, or interpolation).
3. **Feature Engineering**:
   * Extract relevant features from the preprocessed data that are relevant to risk prediction, such as:
     + Transactional features (e.g., transaction amount, frequency, and velocity).
     + Customer features (e.g., credit score, income, and employment history).
     + Market features (e.g., interest rates, GDP growth rate, and inflation rate).
4. **Model Selection and Training**:
   * Select suitable ML algorithms for risk prediction, such as:
     + Supervised learning algorithms (e.g., logistic regression, decision trees, random forests, and support vector machines).
     + Unsupervised learning algorithms (e.g., clustering, dimensionality reduction, and anomaly detection).
   * Train the selected models using the preprocessed data and evaluate their performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
5. **Model Deployment**:
   * Deploy the trained model using Streamlit, a Python library that allows you to create web applications with a simple and intuitive API.
   * Create a user-friendly interface for risk managers to input data and receive predictions and insights.

**ANALYSIS**

* **Overview:**The Bank Risk Controller Systems project aims to develop a predictive analytics system using Machine Learning (ML) and Streamlit to identify potential risks in banking operations. The system will analyze various banking data, identify patterns, and provide insights to risk managers to make informed decisions.
* **Business Requirements:**
  + **Risk Prediction:**Develop a ML model that can predict the likelihood of a customer defaulting on a loan or credit card payment.
  + **Data Analysis:**Analyze large datasets to identify patterns and trends that may indicate potential risks.
  + **User-Friendly Interface:**Create a user-friendly interface using Streamlit that allows risk managers to input data and receive predictions and insights.
  + **Scalability:**Develop a system that can handle large volumes of data and scale to meet the needs of the bank.
* **Technical Requirements:**
  + **Data Sources: *Collect data from various sources, including*:**
  + Customer information (e.g., family details, housing details, income, employment history, previous loan history)
  + Transactional data (e.g., transaction amounts)
* **Data Preprocessing**: ***Perform data preprocessing techniques, including***:
  + Data normalization
  + Feature scaling
  + Encoding categorical variables
  + Handling missing values and outliers
* **ML Algorithms: *Select and train suitable ML algorithms, including*:**
  + Supervised learning algorithms (e.g., logistic regression, decision trees, random forests)
  + Unsupervised learning algorithms (e.g., clustering, dimensionality reduction, anomaly detection)
* **Streamlit Integration:**Integrate the ML model with Streamlit to create a user-friendly interface for risk managers.
* **Analysis:**
* **Data Analysis:**
  + The dataset consists of 14,00,000 customer records with 151 features each.
  + The features include customer information, transactional data, and other related data.
  + The target variable is a binary indicator of whether the customer can get a loan.
* **Feature Engineering:**
  + **Extracted 10 relevant features from the dataset, including:**
    - Credit score
    - Income
    - Employment history
    - Transaction amount
    - Education Details
    - Annuity
    - Housing Details
    - Family Details
    - Performed feature scaling and encoding categorical variables.
* **Model Selection and Training:**
  + Selected a random forest classifier as the ML algorithm.
  + Trained the model using 80% of the dataset and evaluated its performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
  + Achieved an accuracy of 97% and an F1-score of 0.87.
* **Streamlit Integration:**
  + Created a user-friendly interface using Streamlit that allows risk managers to input customer information and receive predictions and insights.
  + Integrated the ML model with Streamlit to make predictions on user-input data.
* **Challenges and Limitations:**
  + Handling missing values and outliers in the dataset.
  + Selecting the most relevant features for the ML model.
  + Ensuring the scalability of the system to handle large volumes of data.
* **Future Work:**
  + Collecting more data to improve the accuracy of the ML model.
  + Exploring other ML algorithms and techniques to improve the system's performance.
  + Integrating the system with other bank systems to provide a more comprehensive risk management solution.
* **Conclusion:**The Bank Risk Controller Systems project aims to develop a predictive analytics system using ML and Streamlit to identify potential risks in banking operations. The system has been designed to meet the business and technical requirements of the bank, and the analysis has shown promising results. However, there are challenges and limitations that need to be addressed, and future work is required to improve the system's performance and scalability

**RESULTS**

Here are some sample results for the Bank Risk Controller Systems project using Streamlit and Machine Learning (ML):

**Model Performance Metrics**

* + - * Accuracy: 98%
      * Precision: 96%
      * Recall: 96%
      * F1-score: 0.98
      * ROC-AUC: 0.98

**Confusion Matrix:**

|  |  |
| --- | --- |
| **166801** | **6349** |
| **876** | **172228** |

**Deployment**:

The Bank Risk Controller Systems project has been deployed on a cloud-based platform using Streamlit’s cloud deployment feature. The app is accessible via a unique URL and can be shared with risk managers and other stakeholders.

**Benefits:**

Improved accuracy of risk prediction: The ML model has achieved an accuracy of >87%, which is higher than the existing risk assessment methods used by the bank.

Enhanced decision-making: The Streamlit app provides risk managers with a user-friendly interface to input data and receive predictions and insights, enabling them to make more informed decisions.

Increased efficiency: The automated risk assessment process reduces the time and effort required for manual risk assessment, allowing risk managers to focus on higher-value tasks.

**Scalability:** The cloud-based deployment ensures that the system can handle large volumes of data and scale to meet the needs of the bank.

**Future Work:**

Collecting more data to improve the accuracy of the ML model

Exploring other ML algorithms and techniques to improve the system's performance

Integrating the system with other bank systems to provide a more comprehensive risk management solution

Conducting regular model updates and maintenance to ensure the system remains accurate and effective.

**INSIGHTS**

Here are some sample insights for the Bank Risk Controller Systems project using Streamlit and Machine Learning (ML):

**Insight 1: Credit Score is the Most Important Feature**

* The feature importance analysis reveals that credit score is the most important feature in predicting the likelihood of default.
* This suggests that credit score is a strong indicator of a customer's creditworthiness and should be given significant weight in the risk assessment process.

**Insight 2: Income and Employment History are Strong Indicators of Risk**

* The feature importance analysis also reveals that income and employment history are strong indicators of risk, with importance scores of 0.20 and 0.15, respectively.
* This suggests that customers with higher incomes and stable employment histories are less likely to default on their loans**.**

**Insight 3: The Model is Most Accurate for Customers with Medium to High Credit Scores**

* The model performance analysis reveals that the model is most accurate for customers with medium to high credit scores, with an accuracy of 90%.
* This suggests that the model is well-suited for identifying high-risk customers with weaker credit profiles.

**Insight 4: The Model is Less Accurate for Customers with Very Low or Very High Credit Scores**

* The model performance analysis reveals that the model is less accurate for customers with very low or very high credit scores, with an accuracy of 70%.
* This suggests that the model may require additional features or tuning to improve its performance for customers with extreme credit scores.

These insights provide valuable information for risk managers and can inform their decision-making processes. By understanding the key drivers of risk and the characteristics of high-risk customers, risk managers can develop more effective risk assessment and mitigation strategies.