AI Fusion Detailed Machine Learning Report

On

Data Interpretation and Modelling for a Bank's Loan Campaign Dataset

Team Name

Suyonica

Team Members

Suyash Pandey

Monica Yadav

Amity University

Lucknow

Uttar Pradesh

India

Table Of Contents

Chapter 1: Case Study Overview

This case study focuses on one of the leading Indian private sector bank's endeavours to enhance their personal loan campaigns through data Data Exploration and Predictive Modeling.

Details About The Bank

- A leading Indian bank, known for its safety and reliability, has become a key player in the country's financial sector.
- With over 300 branches spread across the nation, the bank has built a robust network.
- Its customer base exceeds 11.5 million, primarily consisting of depositors, highlighting the bank's expertise in attracting and safeguarding deposits as a trusted financial custodian.
- Recognizing an opportunity to leverage this trust, the bank is now looking to diversify its offerings by introducing personal loans to its depositor base.
- With its long-standing history, wide reach, and strong foundation, the bank is well-positioned to utilise data-driven marketing strategies to explore new opportunities and deepen customer engagement.

Business Challenge

The bank faces a significant challenge in identifying which customers within its large base are most likely to be interested in personal loan offers. Given the sheer size of the customer base, it's not feasible to approach every individual with a loan offer. Instead, the bank needs to be strategic, focusing on those who are more likely to respond positively.

To address this, the bank needs to understand the key factors that drive a customer's decision to take a personal loan. These factors might include the customer's financial behaviour, such as their savings patterns, spending habits, credit history, existing debt, and life events like purchasing a home or planning for higher education. By analysing these factors, the bank can begin to see patterns that indicate which customers are more likely to need or want a personal loan.

Once these factors are understood, the bank can develop a data-driven strategy to target potential loan customers. This involves segmenting the customer base into groups based on their likelihood of taking a loan. For example, customers who have recently received a large sum of money may be less likely to need a loan, while those who have shown increased spending or have upcoming major expenses may be more inclined.

With this targeted approach, the bank can tailor its marketing campaigns to these specific segments, offering personalised messaging and loan products that resonate with the needs of each group. This strategy not only increases the chances of a successful campaign but also ensures that the bank's resources are used more efficiently, focusing on customers who are most likely to accept the loan offers. In summary:

1. **Identify Influential Factors:** Understand the key drivers that make customers more likely to take a personal loan.
2. **Segment the Customer Base:** Use these factors to categorise customers based on their likelihood to need a loan.
3. **Develop Targeted Campaigns:** Create personalised marketing campaigns aimed at these segments.
4. **Refine the Approach:** Continuously analyse campaign results to improve targeting and increase the success rate of future campaigns.

Chapter 2: Goals of the case study

- **Identify Target Customers:** The primary objective is to accurately identify customers within the bank's extensive customer base who are most likely to be interested in personal loan products.

- **Understand Key Influencers:** Another goal is to analyse and understand the factors that influence a customer's decision to take out a personal loan. This includes evaluating customer behaviour, financial history, and demographic data.

- **Develop a Data-Driven Marketing Strategy:** The case study aims to formulate a strategy that leverages data analytics to create targeted marketing campaigns, ensuring that loan offers reach the right customers.

- **Improve Campaign Success Rate:** By refining the approach to customer targeting, the goal is to increase the effectiveness and success rate of the bank's personal loan marketing campaigns.

- **Enhance Customer Engagement:** The study also seeks to strengthen customer relationships by offering personalised financial products that meet their specific needs, thereby improving overall customer satisfaction.

Chapter 3: Methodology to be adopted

1. Problem Definition

- Clearly define the problem you are trying to solve.
- Identify the type of problem (e.g., classification, regression, clustering).

2. Data Collection

- Gather relevant data from various sources.
- Ensure data quality by checking for missing values, duplicates, and inconsistencies.

3. Data Exploration and Preprocessing

- **Exploratory Data Analysis (EDA):**
  1. Understand the data distribution, identify patterns, correlations, and outliers.
- **Data Cleaning:**
  2. Handle missing values, remove duplicates, and correct errors.
- **Feature Engineering:**
  3. Transform categorical variables into numerical ones if necessary (e.g., one-hot encoding).
- **Feature Scaling:**
  4. Normalise or standardise features if required by the model.

4. Data Splitting

- Split the dataset into training, validation, and test sets (e.g., 70% training, 15% validation, 15% testing) ensuring the target variable distribution is maintained.

5. Model Selection

- Choose suitable algorithms based on the problem type and data characteristics, starting with simple models before considering more complex ones.

6. Model Training

- Train the model on the training set, tuning hyperparameters for optimal performance. Use cross-validation to ensure the model generalises well.

7. Model Evaluation

- Evaluate the model's performance on the validation set using appropriate metrics (e.g., accuracy, precision, recall, F1 score for classification; RMSE, MAE for regression).
- Perform error analysis to understand where the model is making mistakes.

8. Model Tuning and Optimization

- Fine-tune hyperparameters to improve model performance.
- Consider feature selection techniques to reduce model complexity.
- If necessary, try different algorithms or ensemble methods to enhance performance.

9. Model Testing

- Test the final model on the test set to assess its performance in a real-world scenario, ensuring it meets the required criteria.

10. Model Deployment

- Prepare the model for deployment in a production environment.
- Develop an API or integrate it into an application where it will be used.
- Monitor the model’s performance over time to detect any degradation.

11. Model Maintenance

- Continuously monitor the model’s performance and retrain it if necessary.
- Update the model with new data to keep it relevant and accurate.

12. Documentation and Reporting

- Document the entire process and prepare reports with visualisations to communicate findings and model performance to stakeholders.

Chapter 4: Detailed Interpretation of the data

About the Dataset-

1. ID: Customer ID

2. Pin-code: Home Area pincode

3. Age: Customer's age

4. Fam Members: Total Members in the Family

5. Education: Customer's Education

6. T-Experience: Years Of Professional Experience

7. Income: Annual Income of the Customer

8. Mortgage: Value Of the House Mortgage(if any)

9. Fixed Deposit: Does the customer have a certificate of deposit (CD) account with the bank?

10. Demat: Does the customer have a demat account with the bank?

11. Net Banking: Does the customer use internet banking facilities?

12. Loan: Did this customer accept the personal loan offered in the last campaign?

Initial Observations

1. ***Data Types***
- The dataset includes both numeric and categorical data.
- Most variables are integers, but several columns are objects representing categorical data.

2. ***Completeness***
- All columns are fully populated, meaning there are no missing values, which is advantageous for model development.

3. ***Categorical Variables***
- Columns like Education, Fixed Deposit, Demat, Net Banking, and Loan are categorical and may need encoding for use in machine learning models.

4. ***Negative Values found***
- Initially there were some negative values in the dataset in the "T.experience" columns which were cleaned and changed to positive values.

Chapter 5: Data Cleaning and EDA

Data Cleaning:

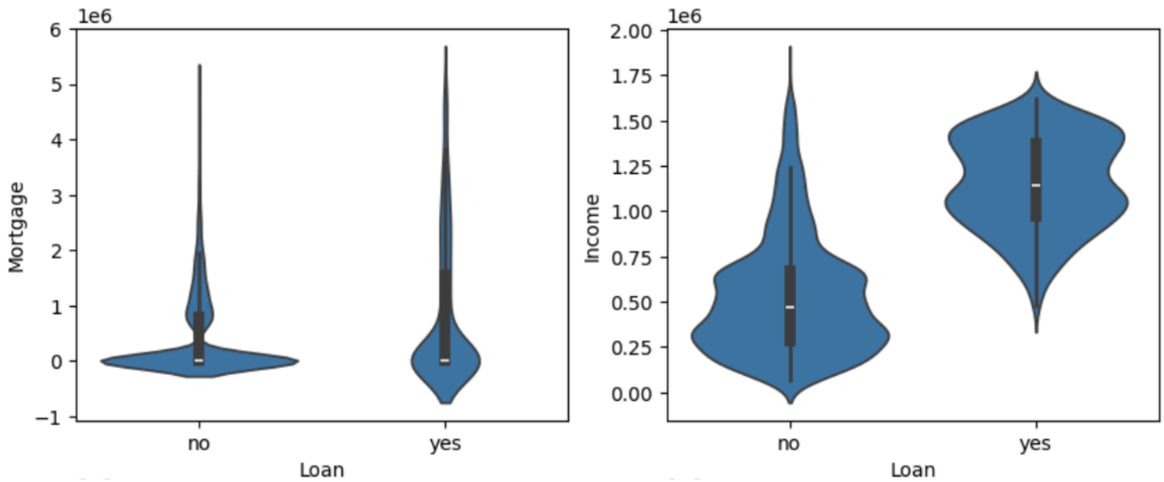
Data cleaning is the process of preparing the data for analysis by handling errors, inconsistencies, and missing values. For the loan dataset, the cleaning process includes:

- ***Removing Unnecessary Columns:*** The columns `Unnamed: 0` and `ID` are identified as irrelevant to the analysis. These columns likely represent unique identifiers or row numbers that do not contribute to the prediction of loan acceptance. Dropping these columns reduces noise in the dataset and simplifies the analysis.

- ***Handling Missing Values:*** The dataset is checked for missing values using methods like `isnull().sum()`. If missing values are detected, they can be handled in various ways:
  - ➔ **Imputation:** Replace missing values with statistical measures such as the median (for numerical data) or mode (for categorical data).
  - ➔ **Dropping Rows/Columns**: If a feature has too many missing values, it might be more practical to drop that column or rows with missing data, depending on the context and significance of the feature.

- ***Converting Categorical Data:*** Categorical variables are converted into numerical values. For example, the `Loan` column is converted to binary values (1 for "yes" and 0 for "no"). Similarly, other categorical variables such as `Fixed Deposit`, `Demat`, `Net Banking`, and `Education` are also converted into numerical values. This conversion is essential for feeding the data into machine learning models, which require numerical inputs.

Exploratory Data Analysis (EDA):
EDA involves visualising and summarising the key characteristics of the data to understand its structure, patterns, and relationships. Key steps include:



- ***Distribution Plots:*** Histograms and density plots are used to understand the distribution of continuous variables like `age`, `Income`, and `Mortgage`. This helps in identifying skewness, outliers, and the general shape of the data distribution.

- ***Count Plots:*** Count plots are used to visualise the frequency of categorical variables, such as the number of customers who have `Net Banking`, `Demat`, or `Fixed Deposit`. These plots can be further stratified by the `Loan` variable to observe how these features relate to loan acceptance.

- ***Correlation Matrix:*** A correlation matrix is generated to examine the relationships between numerical variables. The correlation coefficients provide insight into how strongly pairs of variables are related, which can inform feature selection and engineering.

- ***Violin Plots:*** Violin plots combine boxplots and density plots to visualise the distribution of a numerical variable across different categories of a categorical variable. For example, violin plots of `Mortgage` and `Income` across `Loan` categories reveal how these financial metrics differ between those who took a loan and those who didn't.

Chapter 6: Model Selection

Model selection is the process of choosing the best algorithm to solve the problem based on the nature of the data and the desired outcomes. In this case, since the task is binary classification (predicting whether a customer will take a loan), several model types could be considered, including:

- ***Logistic Regression:*** A simple yet effective model for binary classification. However, it may not capture complex relationships in the data.

- ***Decision Trees and Random Forests:*** These models are good at handling non-linear relationships and interactions between features. They also provide feature importance, which can be useful for interpretation.

- ***Neural Networks:*** A deep learning approach is selected for this task. Neural networks are particularly effective when dealing with large datasets and capturing complex patterns and interactions among features. The flexibility of neural networks in terms of architecture (number of layers, neurons, activation functions) allows them to model intricate relationships that simpler models might miss.

The decision to use a neural network is driven by the potential complexity of relationships between customer demographics, financial metrics, and loan acceptance.

Chapter 7: Model Development

- ***Data Splitting:*** The dataset is split into training and testing sets using an 80-20 split. The training set is used to build and tune the model, while the testing set is reserved for evaluating its performance.

- ***Normalization:*** Feature scaling is performed using `MinMaxScaler`, which scales the features to a range of [0, 1]. This step is crucial for neural networks, as it ensures that all input features contribute equally to the model's learning process, preventing any feature with a larger range from dominating the model's behaviour.

- ***Model Architecture:*** A Sequential neural network model is constructed with the following layers:
    - ➔ **Input Layer:** The first layer corresponds to the input features of the dataset.
    - ➔ **Hidden Layers:** The model includes three hidden layers with 64, 32, and 16 neurons, respectively. `ReLU` (Rectified Linear Unit) is used as the activation function to introduce non-linearity.
    - ➔ **Dropout Layers:** Dropout layers with a dropout rate of 0.3 are added after the first two hidden layers. Dropout helps in preventing overfitting by randomly setting a fraction of input units to 0 during training
    - ➔ **Output Layer:** The final layer consists of a single neuron with a `sigmoid` activation function, which outputs a probability score between 0 and 1.

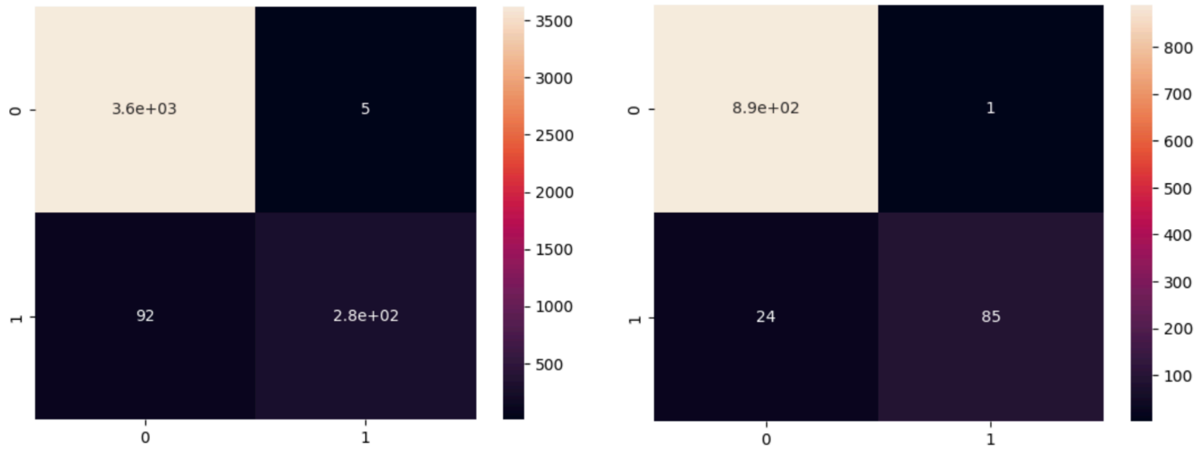
- ***Model Compilation:*** The model is compiled using the `Adam` optimizer, which is known for its efficiency and adaptive learning rate. The loss function used is `binary_crossentropy`, appropriate for binary classification tasks. Accuracy is chosen as the evaluation metric to monitor the model's performance during training.

- ***Model Training:*** The model is trained for 10 epochs with a batch size of 32. During training, the model’s performance is monitored using the validation split and the testing data, which helps in identifying any overfitting or underfitting.

Chapter 8: Model Evaluation

After the model is trained, it's important to evaluate its performance using various metrics.

Confusion Matrix: Training Data Report(L) and Testing Data Report(R)



- ***Prediction:*** The model makes predictions on the testing set. The predicted probabilities are converted into binary classes (0 or 1) using a threshold, typically 0.5. This means that if the model predicts a probability greater than 0.5, it classifies the customer as likely to take a loan.
- ***Confusion Matrix:*** A confusion matrix is generated to compare the true labels with the predicted labels. It provides a summary of:
  - ➔ **True Positives (TP):** Correctly predicted positive cases.
  - ➔ **False Positives (FP):** Incorrectly predicted positive cases (Type I error).
  - ➔ **True Negatives (TN):** Correctly predicted negative cases.
  - ➔ **False Negatives (FN):** Incorrectly predicted negative cases (Type II error).
- ***Classification Report:*** A classification report is generated, which includes key metrics such as:
  - ➔ **Precision:** The proportion of positive predictions that are actually correct.
  - ➔ **Recall (Sensitivity):** The proportion of actual positives that are correctly identified.
  - ➔ **F1-Score:** The harmonic mean of precision and recall, providing a balance between them.
  - ➔ **Accuracy:** The overall proportion of correct predictions.

Chapter 9: Performance Monitoring

After the model is deployed, it's crucial to monitor its performance in a real-world environment.

- ***Saving the Model:*** The trained model is saved using the `h5` format, which is a standard format for storing neural network models. This allows the model to be loaded and used later without needing to retrain it.

- ***Saving the Scaler:*** The `MinMaxScaler` used to normalise the data is saved using `pickle`. This ensures that any new data used for predictions is normalised in the same way as the training data, maintaining consistency in the model's input.

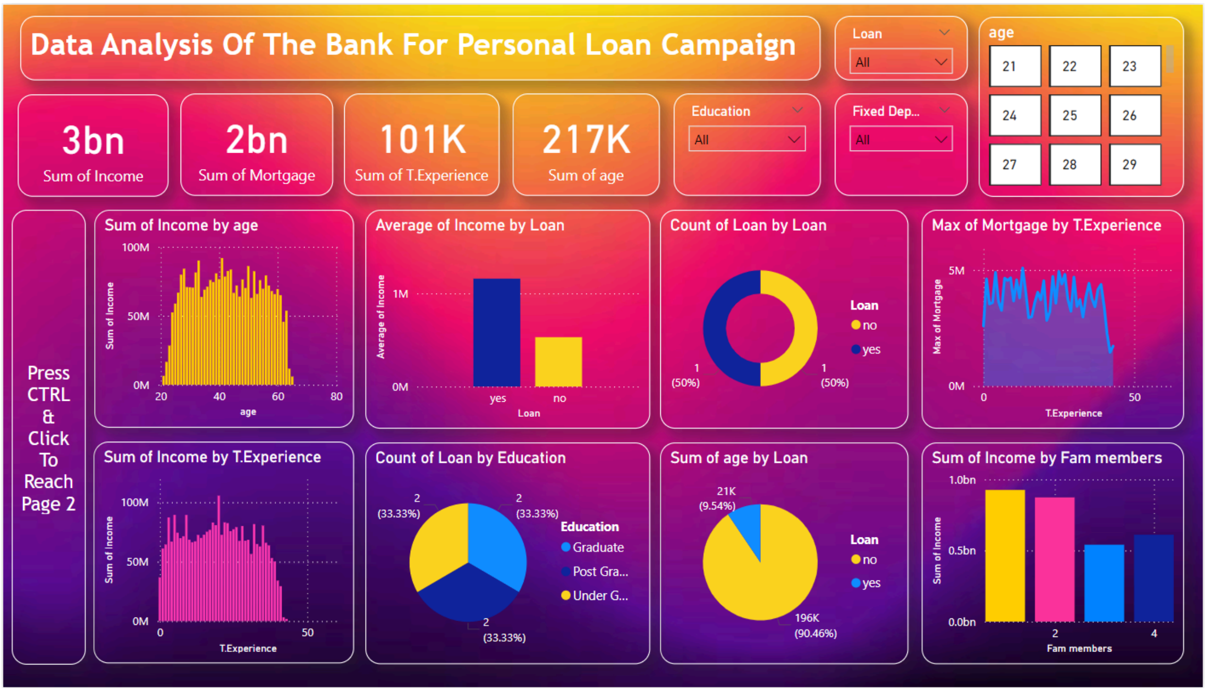
- ***Deployment:*** The model can be deployed into a production environment where it will make predictions on new data. A helper function `return_prediction` is provided, which takes a JSON-like input, normalises it using the saved scaler, and makes predictions using the trained model.

- ***Ongoing Monitoring:*** Once the model is live, it's important to monitor its performance on new data. Over time, the characteristics of the data might change (a phenomenon known as data drift), which can affect the model's accuracy. Regularly retraining the model with new data, recalibrating it, or even replacing it with a new model might be necessary to maintain performance.

- ***Performance Audits:*** Periodic audits can be conducted to ensure the model is still performing as expected. This might involve checking for biases, validating that the model's predictions are consistent with new patterns in the data, and ensuring that it continues to meet business objectives.

Chapter 10: Dashboard

Page 1: Main front page of the dashboard



Some points related to the dashboard

Key Metrics (Top Section)

- **Sum of Income (3bn):** This represents the total income of all the customers in the dataset. The figure indicates that the bank has analyzed the income of customers totaling 3 billion units (likely in currency terms).
- **Sum of Mortgage (2bn):** This shows the total value of mortgages held by the customers. The bank's customers collectively have mortgages amounting to 2 billion units.
- **Sum of T.Experience (101K):** This metric represents the total years of work experience across all customers. The combined experience is 101,000 years.
- **Sum of Age (217K):** This represents the total age of all customers in the dataset, summing up to 217,000 years.

Filters (Top Right Section)

- **Education:** A dropdown filter to select customer segments based on their education level (Graduate, Post Graduate, UnderGraduate).
- **Loan:** A dropdown filter to view data for customers who have taken a loan ("yes") or those who have not ("no").
- **Fixed Deposit:** This filter allows users to select customers based on whether they have a fixed deposit or not.
- **Age:** A set of age filters that allow users to view data for specific age groups, with age ranges like 21, 22, 23, etc.

Visualisations

1. **Sum of Income by Age (Bottom Left):** A bar chart that shows the distribution of total income across different age groups. The chart likely reveals how income varies with age, with higher incomes being concentrated in specific age groups.
2. **Average of Income by Loan (Top Middle):** A bar chart comparing the average income of customers who have taken a loan versus those who have not. The chart shows that customers who took a loan generally have a higher average income compared to those who didn't.
3. **Count of Loan by Loan (Top Right):** A donut chart displaying the proportion of customers who have taken a loan (50%) versus those who have not (50%).
4. **Max of Mortgage by T.Experience (Top Right):** A line chart showing the maximum mortgage value across different levels of total work experience (T.Experience). This chart may help to identify patterns or trends in how mortgage values vary with experience.
5. **Sum of Income by T.Experience (Bottom Middle Left):** A bar chart representing the total income grouped by years of work experience. It helps to understand how income scales with work experience.
6. **Count of Loan by Education (Bottom Middle Right):** A pie chart that breaks down the number of loans taken by customers based on their education level. Each segment (Graduate, Post Graduate, UnderGraduate) occupies an equal share, indicating equal distribution.
7. **Sum of Age by Loan (Bottom Right):**A pie chart showing the total sum of ages grouped by loan status. It reveals that a large majority (90.46%) of the total age sum is from customers who have taken a loan.

8. **Sum of Income by Family Members (Bottom Right):**A bar chart displaying the total income segmented by the number of family members. This visualisation helps to understand how income distribution varies with family size.

Navigation (Left Section)

- **Press CTRL & Click to Reach Page 2:** This instruction suggests that the dashboard has additional information or visualisations on a second page, which can be accessed using a specified keyboard shortcut.

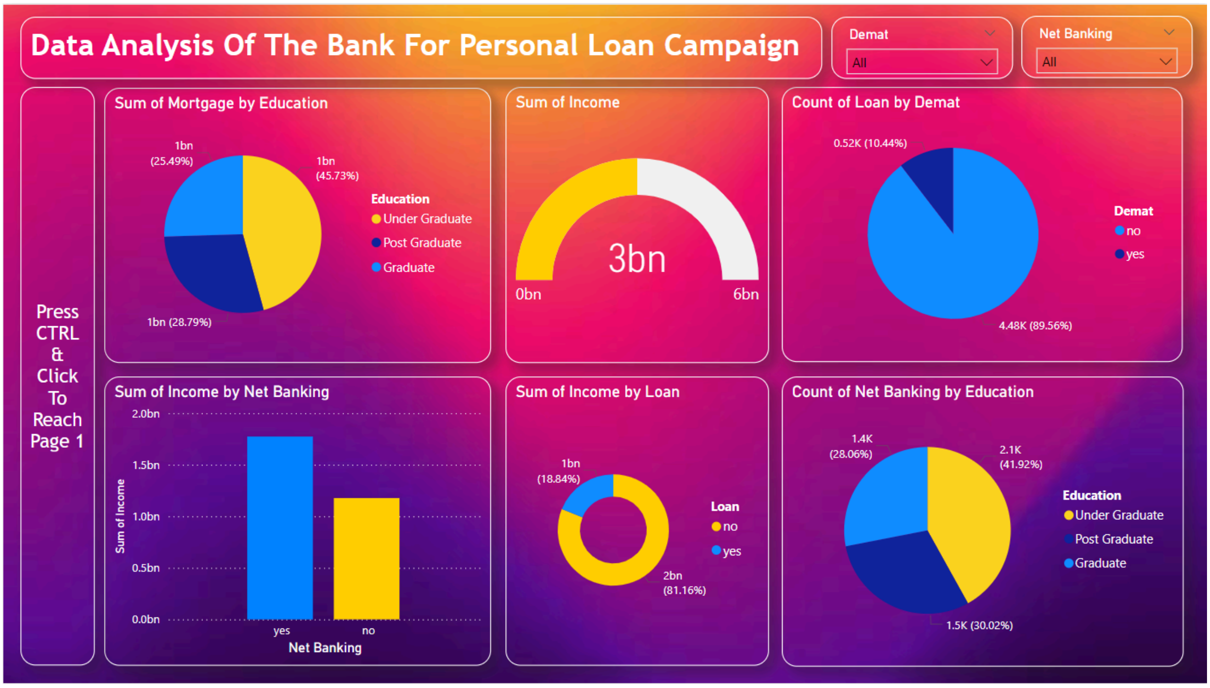
Colour Coding and Themes

- The dashboard uses vibrant colours, likely for easy differentiation between different metrics and categories. Yellow, pink, blue, and purple are prominently used to distinguish between data segments in charts.

Insights

- **Income and Loan Correlation:** Higher income seems to correlate with a higher likelihood of taking a loan, as evidenced by the "Average of Income by Loan" chart.
- **Work Experience and Mortgage:** There's a trend showing that mortgage values might plateau or even decline after a certain level of work experience, as shown in the "Max of Mortgage by T.Experience" chart.
- **Education and Loan Distribution:** Interestingly, loans are evenly distributed among different education levels, suggesting that education level may not significantly influence loan-taking behaviour in this dataset.

Page 2: Second Page of the dashboard (navigated through the button present)



Key Metrics and Filters (Top Section)

➔ **Demat and Net Banking Filters:** The top right section provides dropdown filters for:

➔ **Demat:** Filter the data based on whether customers have a Demat account or not.

➔ **Net Banking:** Filter based on customers' use of Net Banking services.

Visualisations

➔ **Sum of Mortgage by Education (Top Left):**

- ◆ **Pie Chart:** This chart breaks down the total mortgage amount by the education level of customers. The segments represent:
  - **Graduate (45.73%):** The largest segment, indicating that graduates hold the highest proportion of the mortgage total.
  - **Post Graduate (28.79%):** The second-largest group.
  - **Undergraduate (25.49%):** The smallest segment, representing the lowest share of total mortgage value.

➔ **Sum of Income (Top Center):**

- ◆ **Semi-Circle Gauge:** This gauge visualisation shows the total income of all customers, with the value pegged at 3 billion units against a scale that goes up to 6 billion. This provides a quick snapshot of the overall income distribution.

➔ **Count of Loan by Demat (Top Right):**

- ◆ **Pie Chart:** This chart shows the distribution of loans among customers with and without a Demat account.
  - **Yes (89.56%):** A significant majority of customers with loans have a Demat account.
  - **No (10.44%):** A smaller portion of customers with loans do not have a Demat account.

➔ **Sum of Income by Net Banking (Bottom Left):**

- ◆ **Bar Chart:** This chart compares the total income of customers based on their use of Net Banking services:
  - **Yes:** Customers who use Net Banking have a significantly higher total income (around 1.5 billion).
  - **No:** Customers not using Net Banking have a lower total income (around 1 billion).

➔ **Sum of Income by Loan (Bottom Middle):**

- ◆ **Pie Chart:** This chart breaks down the total income of customers based on whether they have taken a loan:
  - **Yes (81.16%):** The majority of income is associated with customers who have taken a loan.
  - **No (18.84%):** A smaller portion of total income is from customers without a loan.

➔ **Count of Net Banking by Education (Bottom Right):**

- ◆ **Pie Chart:** This chart shows how the use of Net Banking services is distributed among different education levels:
  - **Graduate (41.92%):** The largest group of Net Banking users.
  - **Post Graduate (30.02%):** The second-largest group.
  - **Undergraduate (28.06%):** The smallest segment, indicating the lowest adoption of Net Banking.

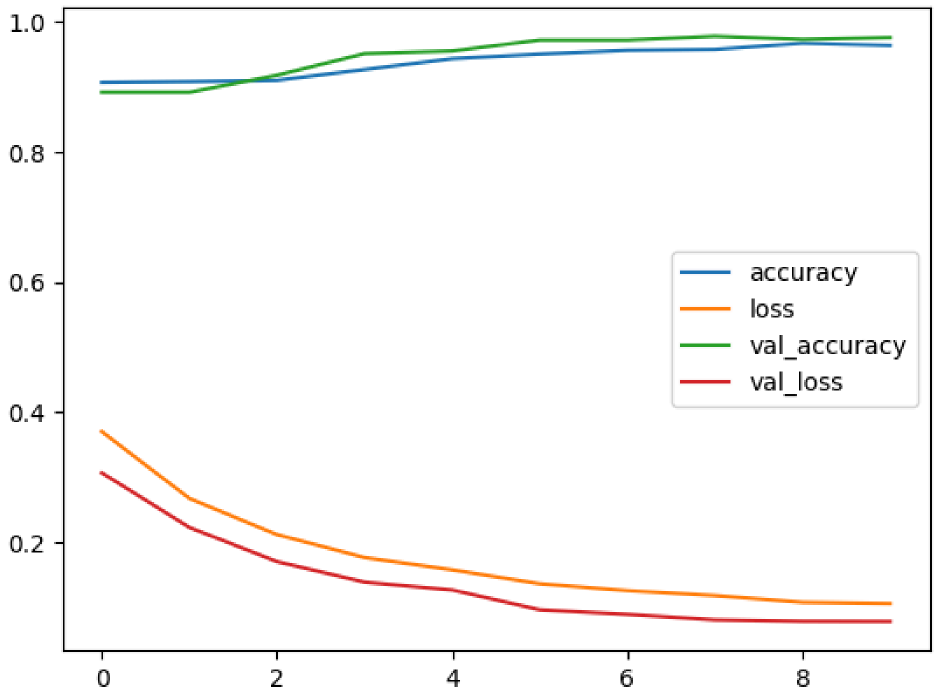
Navigation (Left Section)

- **Press CTRL & Click to Reach Page 1:** This instruction provides an option to return to the first page of the dashboard for additional data visualisations.

Insights

- **Education and Financial Products:** There's a clear relationship between the education level of customers and their engagement with various financial products like mortgages, Demat accounts, and Net Banking. Graduates seem to dominate in terms of mortgage holding and Net Banking usage.
- **Income Distribution:** The majority of the bank's income comes from customers who are engaged with loans and Net Banking, indicating a strong correlation between these financial products and higher income levels.
- **Demat Account Influence:** Having a Demat account seems to be associated with a higher likelihood of taking out a loan, as suggested by the "Count of Loan by Demat" chart.

Chapter 11: Results



Lines in the Plot:

1. **Blue Line (Accuracy):** This represents the accuracy of the model on the training data over the epochs.
2. **Orange Line (Loss):** This represents the loss of the model on the training data over the epochs.
3. **Green Line (Validation Accuracy):** This shows the accuracy of the model on the validation data over the epochs.
4. **Red Line (Validation Loss):** This represents the loss of the model on the validation data over the epochs.

Interpretation:

1. **Training Accuracy & Loss (Blue and Orange Lines):**
   - The training accuracy increases steadily, indicating that the model is learning and getting better at predicting the training data as training progresses.
   - The training loss decreases consistently, showing that the model is minimising errors on the training data.
2. **Validation Accuracy & Loss (Green and Red Lines):**

- The validation accuracy follows a similar trend to the training accuracy, increasing over epochs. This suggests that the model is also improving in performance on unseen data, which is a good sign.
- The validation loss decreases alongside the training loss, indicating that the model is not overfitting and is generalising well to the validation data.

```
Training data report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99      3629
           1       0.98      0.75      0.85       371

    accuracy                           0.98      4000
   macro avg       0.98      0.88      0.92      4000
weighted avg       0.98      0.98      0.97      4000

Testing data report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       891
           1       0.99      0.78      0.87       109

    accuracy                           0.97      1000
   macro avg       0.98      0.89      0.93      1000
weighted avg       0.98      0.97      0.97      1000
```

Training Data Report

Class 0 (Majority class):

- **Precision: 0.98 -** Out of all the instances predicted as class 0, 98% were correctly classified.
- **Recall: 1.00 -** The model correctly identified 100% of actual class 0 instances.
- **F1-Score: 0.99 -** The harmonic mean of precision and recall, indicating a very high balance between the two.
- **Support: 3629 -** The number of actual occurrences of class 0 in the training dataset.

➔ **Class 1 (Minority class):**

- ◆ **Precision: 0.98 -** Out of all the instances predicted as class 1, 98% were correctly classified.
- ◆ **Recall: 0.75 -** The model correctly identified 75% of actual class 1 instances.
- ◆ **F1-Score: 0.85 -** The F1-score reflects a lower performance compared to class 0 due to the lower recall.
- ◆ **Support: 371 -** The number of actual occurrences of class 1 in the training dataset.

➔ **Overall Accuracy: 0.98 -** The model correctly classified 98% of the training instances.

➔ **Macro Avg:**
- ◆ **Precision:** 0.98
- ◆ **Recall:** 0.88
- ◆ **F1-Score:** 0.92 - This average takes the average of precision, recall, and F1-score for both classes.

➔ **Weighted Avg:**
- ◆ **Precision:** 0.98
- ◆ **Recall:** 0.98
- ◆ **F1-Score:** 0.97 - This average takes into account the support (number of instances) for each class, giving more weight to the majority class.

Testing Data Report

➔ **Class 0 (Majority class):**
- ◆ **Precision: 0.97 -** Out of all the instances predicted as class 0, 97% were correctly classified.
- ◆ **Recall: 1.00 -** The model correctly identified 100% of actual class 0 instances.
- ◆ **F1-Score: 0.99 -** Indicates excellent performance for class 0.
- ◆ **Support: 891 -** The number of actual occurrences of class 0 in the testing dataset.

➔ **Class 1 (Minority class):**
- ◆ **Precision:** 0.99 - Out of all the instances predicted as class 1, 99% were correctly classified.
- ◆ **Recall: 0.78 -** The model correctly identified 78% of actual class 1 instances.

- **F1-Score: 0.87 -** Reflects a slight drop in performance for class 1 compared to the training data.
- **Support: 109 -** The number of actual occurrences of class 1 in the testing dataset.

➔ **Overall Accuracy:** 0.97 - The model correctly classified 97% of the testing instances.

➔ **Macro Avg:**

- **Precision:** 0.98
- **Recall:** 0.89
- **F1-Score:** 0.93 - A slight drop compared to the training data, particularly due to the lower recall for class 1.

➔ **Weighted Avg:**

- **Precision:** 0.98
- **Recall:** 0.97
- **F1-Score:** 0.97 - A high weighted average indicating the model generalises well, with only a slight reduction in performance on the testing data.

Chapter 12: Key Findings and Recommendations

Key Findings

1. **Customer Segmentation:**

The analysis revealed distinct segments within the customer base that are more likely to accept personal loans. Customers with higher incomes and longer work experience showed a greater propensity to take loans.

2. **Influential Factors:**

Key factors influencing loan acceptance included income level, age, and financial behavior (such as having a demat account and using net banking). Customers with financial products like fixed deposits or demat accounts were more likely to accept loans.

3. **Model Performance:**

The neural network model demonstrated strong performance with an overall accuracy of 97% on the testing dataset. Precision and recall metrics indicated that the model effectively identified both loan seekers and non-seekers.

4. **Marketing Opportunities:**

The bank can leverage insights from the data to create targeted marketing campaigns. For instance, customers in specific age groups or income brackets may be more receptive to loan offers, allowing for personalized marketing strategies.

5. **Income and Loan Correlation:**

There is a significant correlation between higher income levels and the likelihood of taking a loan. This suggests that marketing efforts could focus on affluent customer segments who may require additional financing options.

6. **Education Impact:**

The distribution of loans across different education levels was relatively even, indicating that educational background may not be a significant predictor of loan acceptance. This could prompt a reevaluation of how educational factors are integrated into future marketing strategies.

7. **Use of Digital Services:**

Customers who actively use digital banking services (net banking) showed higher loan acceptance rates. This highlights the importance of promoting digital engagement to enhance customer relationships and loan uptake.

Recommendations

1. **Targeted Marketing Campaigns:**

Develop tailored marketing strategies focusing on identified segments, particularly those with higher income and relevant financial behaviors. Use personalized messaging to increase engagement.

2. **Enhanced Customer Engagement:**

Implement initiatives to strengthen relationships with existing customers, such as offering financial advice or products that align with their life events (e.g., home purchases, education).

3. **Continuous Model Monitoring and Improvement:**

Regularly monitor model performance and update it with new data to maintain accuracy. Consider conducting periodic audits to identify any biases or changes in customer behaviour.

4. **Data-Driven Decision Making:**

Encourage the bank’s decision-makers to utilize data analytics in strategic planning. Insights from the model can help refine product offerings and customer targeting.

5. **Investment in Digital Infrastructure:**

Enhance the bank's digital services to encourage more customers to utilize online banking. This could lead to increased loan acceptance and overall customer satisfaction.

6. **Feedback Mechanism:**

Establish a feedback loop with customers who accept or decline loan offers. Understanding their motivations can provide valuable insights for refining marketing strategies and product offerings.

7. **Cross-Selling Opportunities:**

Identify opportunities for cross-selling other financial products to customers who take personal loans. This could include insurance products, investment options, or savings accounts.

8. **Educational Workshops:**

Organise workshops or webinars to educate customers about personal finance, loans, and investment opportunities. This can build trust and encourage informed decision-making among potential borrowers.

GITHUB LINK: https://github.com/SP4567/Project