IBA Institute of
Business Administration
Karachi
Leadership and Ideas for Tomorrow

IBA SMCS
School of Mathematics and Computer Science

<u>INTELLISCORE - Ai Based Credit Scoring System</u>

This BS Project report is submitted to the Department of Computer Science as partial
fulfillment of Bachelor of Science in Computer Science degree

https://github.com/Fazulsden/Fyp-SRS.git

by

Shahmeer Khan, Fazul Al Rehman. Nosherwan Sheikh, Abdur Rehman

Advised by

Miss Tasbiha Fatima
Lecturer
Department of Computer Science
School of Mathematics and Computer Science (SMCS)
Institute of Business Administration (IBA) Karachi

Fall 7th semester 2024
Institute of Business Administration (IBA) Karachi Pakistan

Contents

Chapter 1

Problem Statement

This chapter provides an overview of significance and methodology of this research. It also provides a brief outline of the flow of the whole dissertation and the overview of what is to be covered.

1.1 Case overview

Credit scoring is an integral part of the banking system and its traditional trusted methods are slowly but surely leading to inefficiency as time goes on in the age of automation, this further leads to significant drawbacks for both financial institutions and their clients. These approaches typically involve manual calculations, hence requiring banks to allocate significant resources and personnel to the task of score computation, which can prove to be both time-consuming as well as financially extensive. Due to this scalability also becomes and issue in the long run as an increasing number of clients can cause further blocks in the pipeline of operations that follow credit score calculation. This is the bank's perspective at this point.

From the clients' perspective, traditional credit-scoring methods are undoubtedly unreliable due to their dependance on facts and figures. These methods often fail to account for the nature of an individual's financial behavior. For instance, they typically consider only a limited number of parameters, such as income, debts, and repayment history, without properly taking into account their relative importances. Banks fail to factor in crucial information like the client's health, their dependent family members, spending habits etc as these add to the already extensive computation for credit score calculation. As a result, clients frequently receive scores that are not representative of the actual situation which makes them lose trust in banks that handle their scoring in such a way. This lack of accuracy can lead to unfair assessments, with some clients being unfairly scrutinized and given low scores thus missing out on beneficial credit oppurtunites, which is detrimental for both client and bank. Furthermore, this isn't a case of inability but more so of ignorance. Traditional systems lack the ability to analyze the diverse

data sources available in today's digital age. With the increasing digitization of financial trans-
actions and the availability of alternative data sources, such as spending patterns, utility bill
payments, and social behaviors, it is evident that a more comprehensive and data-driven ap-
proach is necessary to ensure fairness and accuracy in credit scoring. A complete integration
of all possible factors is crucial to accurately assess an individual's credit score.

The solution to this is "INTELLISCORE", an automated AI-based approach which presents
a transformative solution to these challenges. By leveraging advanced machine learning algo-
rithms, the credit-scoring process can transition from being formula-driven to data-driven. This
approach enables the integration of thousands of records and dozens of features into the scoring
process, capturing the nuances and complexities of individual financial profiles. Unlike tradi-
tional methods, AI-based systems can identify patterns and correlations within the data that
may not be immediately apparent through manual analysis. Where traditional systems fail to
capture the correlation between several different factors that influence each other's contribution
to the final score, Intelliscore leverages the new era technology available to us to accurately as-
sess all that could contribute to an individual's score, figure out complex inter relations as well
as capture which factors hold the most importance among others.

For banks, the adoption of an AI-based credit-scoring system offers substantial benefits. The
automation of the scoring process eliminates the need for extensive manual labor, reducing
operational costs and enabling the reallocation of resources to more strategic initiatives. Ad-
ditionally, AI models can process applications at a much faster rate, improving efficiency and
enhancing customer satisfaction by delivering quicker decisions. The bank staff accessing the
system won't need a complex understanding of how machine learning models as the solution
for them will be only a click away.

On the clients' side, our credit-scoring system ensures a more reliable evaluation of creditwor-
thiness. By considering a broader range of data points and dynamically adjusting the weight
of each parameter based on its significance, these systems provide scores that more accurately
reflect individual circumstances. This approach not only reduces the likelihood of unfair as-
sessments but also opens up access to credit for individuals who may have been overlooked
by traditional methods. Thus, further enabling deeper trust between the client and the bank
ensuring everyone leaves satisfied.

In conclusion, the transition from traditional credit-scoring methods to an automated AI-based
approach addresses the core inefficiencies and inaccuracies that have long affected the finance
industry. By harnessing the power of AI, banks can streamline operations, reduce costs, and im-
prove scalability, while clients benefit from fairer, data-driven evaluations that align with their
unique financial profiles. This technological shift is essential to modernizing credit-scoring.

1.2   Scope

The scope of this project covers the end-to-end development of the AI-based credit scoring system. It will encompass three key components: machine learning for predictive modeling, a full-stack web-based platform for user interaction, and a data dashboard for real-time monitoring and reporting. This will streamline the lending process by offering faster and more accurate risk assessments while ensuring scalability for financial institutions operating at various levels. It includes data collection and pre-processing, model design and training, development of a web-based UI for user interactions, creation of interactive dashboards, and the integration of data security and compliance protocols. This system will be developed to support scalability, allowing multiple banks/financial institutions to adopt it, and will be adaptable to different regulatory environments. The project will also focus on reducing bias in credit scoring to promote fairness.

Chapter 2

System Requirements

This portion will feature both the functional, non functional requirements as well as the features that we are looking to implement i.e. user interface requirements.

2.1 Functional Requirements:

Detailed specifications of the essential operations and tasks the system must perform to achieve its objectives.

We will be disclosing the basic functionality requirements that so far our team has recognized. Please note that further information may or may not cause us to remove some functionality or not. Right now, these functionalities are the basic and important functionalities alongside other requirements that we would be implementing. Their level could be advanced as well, and this applies to both functional as well as nonfunctional requirements! Further, if we see any changes, we would let our mentor Ma'am Tasbiha Fatima and the FYP Committee know through the use of resubmission or in any further document.

P.S These are basic functional requirements; hence, no priority has been assigned to them!

Functional Requirements	Description
User Registration & Authentication	The system must support user registration and authentication via usernames, passwords, and multi-factor authentication (MFA) to ensure only authorized bank employees or loan officers can access sensitive financial data. Each user will have a unique account linked to their role for access control and data integrity.
Data Input for Credit Assessment	Users must be able to input relevant financial data for credit assessments (Reports). This includes personal information (name, CNIC, address) and financial details like income, expenses, existing liabilities, and credit history. More importantly Meezan bank would already have their existing data of the customers so that would be used on its when employees input information
AI-Based Credit Scoring (MODEL)	The Model would be using AI-based algorithms, including **Machine Learning, NLP(GENAI) And Computer Vision**: The Reason would be obvious for Scanning Documents making Sure its up to date/Real, and further ML /NLP to assess the creditworthiness of users based on the data
	entered. They will analyze factors like income, credit history, and liabilities to generate a **credit score**. The system will be trained on historical data specific to Customers credit behaviors and financial conditions. The Model will learn to adapt based on the different patterns in credit applications to provide accurate assessments that align with the bank's criteria.
Credit Report Generation	After the Model processes the data, the system will generate a comprehensive credit report. This report will detail the user's credit score, factors affecting it (e.g., timely bill payments, outstanding loans), and a set of suggested actions (e.g., improve credit utilization, pay off high-interest loans). The report will be in a format (PDF/Excel) that can be easily shared or printed for bank employees to review during loan assessments.

User Profile Management (Optional)	Right now, we have been not entertained with the data (if we would or would not given access to the public) so we do not know weather or not we would be allocating the Dashboard to the public instead our current goal is that the bank employee may have the Access This feature would be redundant for him In some Cases but I am adding since we might give the access to the public!
Decision-Making Logic	Based on the generated credit score, the system should provide a decision support system that indicates whether a user qualifies for a loan, credit card, or other financial products. This logic will consider predefined rules and thresholds set by the bank, such as a minimum credit score required for approval. The system could also suggest the maximum loan amount or credit limit based on the user's financial profile.
Notifications	The system should notify users (or bank employees) about any significant changes to a user's credit score. Also, users should be notified of upcoming loan repayments, opportunities for improving their score, or required actions (e.g., submitting updated financial documents). The notifications will be sent via email or SMS, especially for Our Countries Customer users who may prefer these channels.
Credit Score History	The system should be capable of showing a history of the credit score for a time period, thereby enabling the user and bank employees to trace changes over time as well as determine trends. This feature would be of immense use for loan officers analyzing long-term behavior and also for users working on an improvement in credit score.

2.2   Non-Functional Requirements:

Standards and constraints that define the quality attributes, performance, and reliability of the system. Added a Disclaimer in Functional Requirements due to the neatness of the report I am
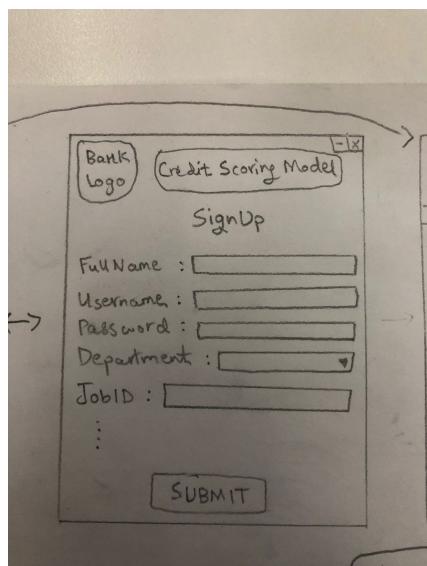
not entering it again here. Kindly Please follow the Pattern of the report so you may get the gist of the idea how have we functioned without proper replies and engagement from the side of our industry Heads!

| Non-Functional Requirement	Description	Priority
High Availability	The system needs to ensure 99.9% availability to support continuous access by bank employees, particularly in high-demand environments. As the system is to be used internally within the bank, any downtime could severely affect loan assessments and processing.	High
Performance	The system should be able to process credit score calculations and report generations within Minimal (As low as few Seconds) seconds per request. Fast processing is important because loan officers need to assess and decide without delay, especially at peak times when a number of loan applications are processed.	High
Scalability (incase of giving the access to the public)	The system should be designed to scale as the bank expands and the number of users grows. This includes handling larger volumes of financial data, accommodating increasing numbers of customers, and supporting the addition of new bank branches or services.	Medium
Data Privacy & Security	Given the sensitive nature of financial data, the system must ensure high levels of security and data privacy. This includes encryption of personal and financial information both in transit and at rest. The system must comply with local data protection regulations in Pakistan, ensuring that personal financial data is not compromised. Access control mechanisms will also restrict access to financial information based on user roles. Good thing is that Pakistani banks have the highest of the quality in the Data Security Section hence we could potentially integrate this within the bank's application!	High
Integration with External Data Sources	It must integrate with external credit bureaus or databases such as the State Bank of Pakitan or Pakistan Credit Bureau or other financial data sources that will fetch real-time credit data. This will result in more accurate and updated assessments, especially when	High
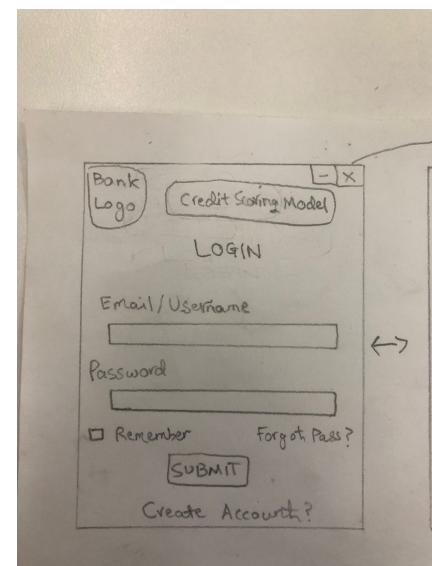
|		
	applicants have a history with other financial institutions. (There are Services used by the SBP That show the exact history of the client since local banks usually don't share details of the customer with each other hence SBP Acts as Sort of Middle man for most of the criteria's/Services.	
Mobile Compatibility (Incase of our Model being open to The Public)	While this system is primarily used internally by bank employees, mobile compatibility is important for remote access by bank officers who may need to assess credit scores while out in the field. The system should be optimized for mobile devices (smartphones, tablets) to ensure that it remains accessible and functional across all platforms.	Medium
Internationalization	Though the system would be specific to Pakistan, it could potentially be multi-language and different currency-based, which might help a bank with international clients or with plans to expand to other regions in the future.	Low

2.3   User Interface Requirements:
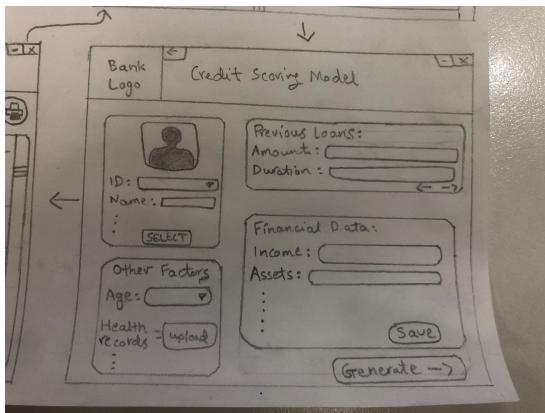
Specifications for the design and usability features that ensure an intuitive and user-friendly interaction with the system.
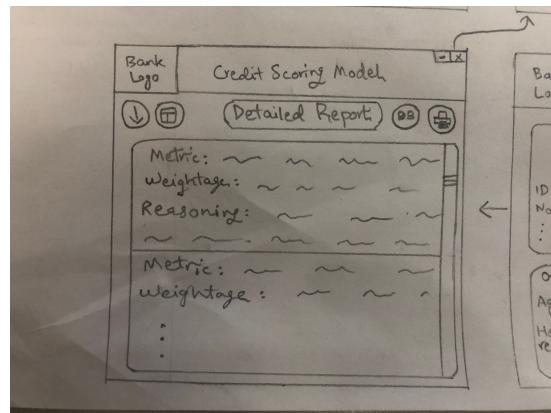

Signup Screen
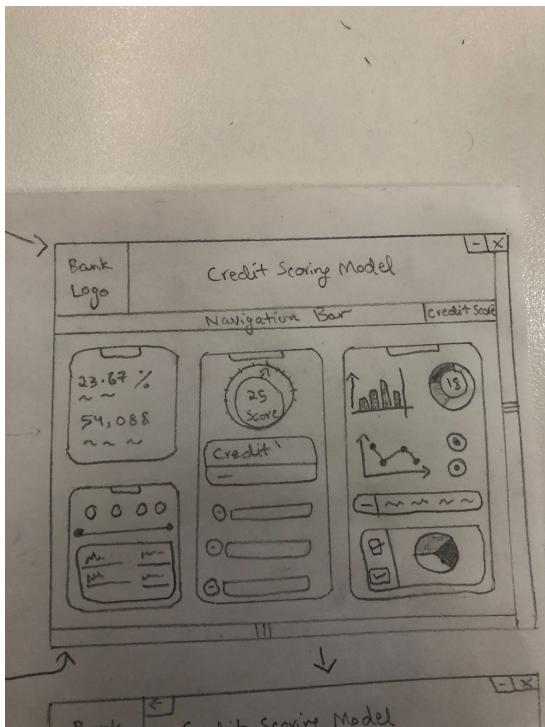
Login Screen

**Detail entry Screen**
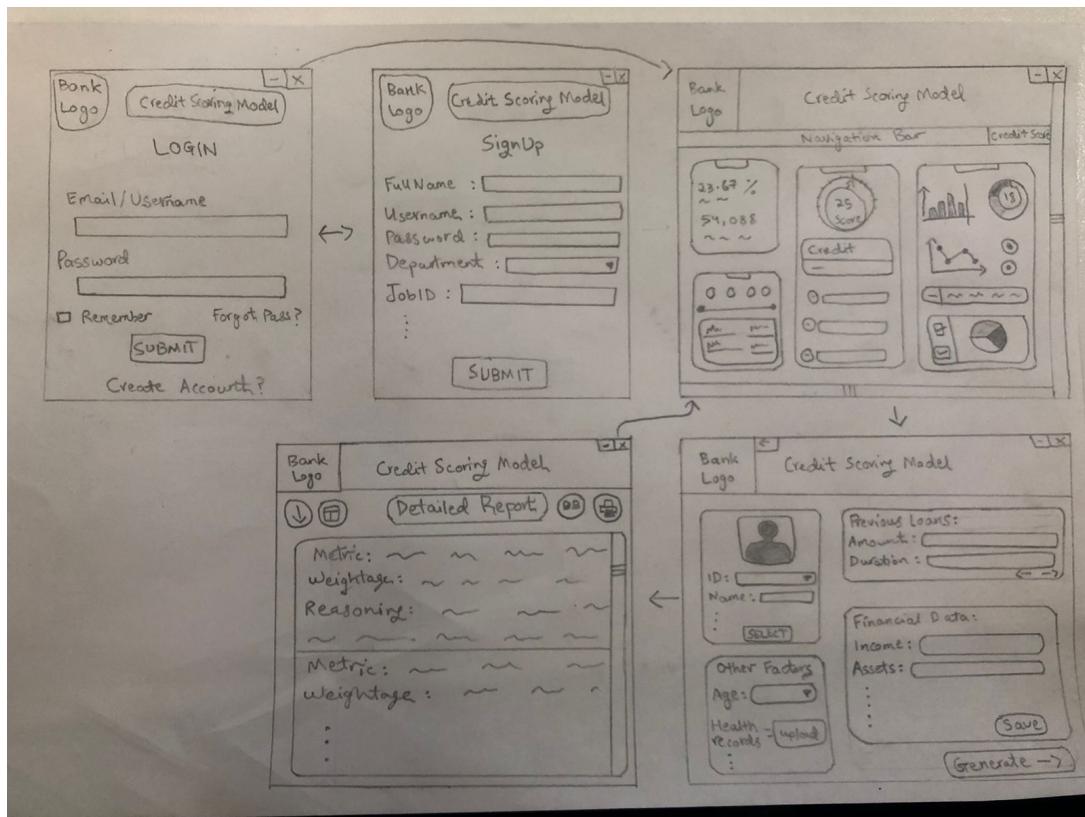
Report Screen

Dashboard Screen

2.3.1 Complete overview/process flow

This describes a complete overview of a potential user journey

1. The user starts off on the login screen. If an account exists, the user is then taken to the home/dashboard screen.

2. If an account doesn't exist, the user can go to the signup page using "Create Account" to create an account and log in using that.

3. At the home screen, aside from other options on the navigation screen, the user opts for the credit score option. The home page will feature random sample dashboards as examples.

4. Upon choosing this, the user is taken to the details page, where they are prompted to fill in the relevant details to determine the credit score.

5. From here, the user is taken to the report page, which features an in-depth analysis. The user can choose to download, print, or save the report.

6. The user can then go to the dashboard page from here, where they can view the generated dashboards, relevant to the recently generated credit score.

Chapter 3

Functional Requirements Specification

Comprehensive outline of the system's functionalities, including use cases that align with stakeholder objectives and the goals of an AI-driven credit scoring system.

3.1 Stakeholders

Right now, so far whatever we have founded/Gathered show us that these would be the Stake Holders along with their Responsibilities!

Stakeholder	Role/Responsibility
End Users (Customers)	Individuals who provide their financial information (income, debts, employment status) and receive a credit score. They may also view and monitor their credit history, Possible Manage (Incase).
Loan Officers / Loan Approval Committee	Employees (or Committee) from the bank who assess applicants for loan eligibility based on AI-generated credit scores and recommendations. They use our system to evaluate if a user qualifies for a loan or credit card.
Frontend/Backend Developers	Full-Stack Developer Looking After Everything (Front end/ Backend Potentially database Along side with Collaborating with all the other members specially with the Model Maker)
AI Model Developers (our Team)	Data scientists, engineers who develop and train the machine learning models that generate credit scores based on users' data. They ensure that the Model Would be optimized for local financial behaviors in Pakistan. Especially Meezan Bank
System Administrators (Maintenance Crew)	These include the internal IT staff, which are supposed to manage the system configurations, manage access, monitor performances, and ensure that the system is stable and secure. They will make sure the access of the system is appropriately restricted to the authorized bank employees.
Business Analysts	Professionals who gather and refine the business requirements to ensure that the system meets the bank's needs for loan processing, credit scoring, and compliance with local regulations.
Compliance Officers	Individuals responsible for ensuring that the system adheres to data protection laws (SBP: - State Bank Of Pakistan's LAW's/ Regulations) GDPR, and other regulations such as financial compliance and anti-money laundering (AML) standards. They ensure that sensitive data is handled appropriately.

3.2   Actors and Goals

We would have followed the Table format but since Already in 2.1 , 2.2 and 3.1 we have already used Tables and due to high Description Sizes they look oversized so for this part we would be focusing on more brief approach!

1. **End Users (Customers)**

   Goal: To view, monitor, and improve their credit score based on personal financial data (income, debts, employment). They aim to track their credit score, receive improvement suggestions, and check eligibility for loans or credit products.

2. **Loan Officers (Bank Employees) / Loan Approval Committee**

   Goal: To assess loan eligibility using the AI-generated credit scores. And Manual If Required, Loan officers will make decisions on loan approvals or rejections, and advise customers on improving their credit scores if necessary.

3. **Full Stack Developer**

   Goal: Responsible for handling everything from the front end to the backend of both the applications App/Web, Also Database handling as well, the actor would closely work with Model Developer!

4. **AI Model (System) / Developers**

   Goal: To process user data and generate accurate credit scores using machine learning. The model also provides recommendations to improve credit scores and adapts over time based on new data and financial behaviors in Pakistan. Improvement to the model would be done by the Developers.

5. **System Administrators (IT Team)**

   Goal: To ensure system stability, security, and performance. Administrators manage user access, monitor system health, and perform regular maintenance to keep the system running smoothly. They Also Maintain everything!

6. **Business Analysts**

   Goal: To gather and define business requirements, ensuring the system aligns with the bank's goals for loan processing, credit scoring, and compliance with regulations. They work closely with other stakeholders to refine system features. So that at the end of the day the accuracy/predictions of the model should be more accurate!

7. **Compliance Officers**

   Goal: To ensure the system adheres to data protection and financial regulations (e.g., SBP laws in Pakistan, GDPR). They ensure secure handling of sensitive data and regulatory compliance, minimizing risks related to privacy and financial laws.

3.3 Use Cases

Detailed representation of interactions within the system, focusing on critical processes like data ingestion, model training, score generation, and result interpretation.

Use Case Name	Description	Actors	Preconditions	Main Flow	Postconditions
Credit Score Generation	The system calculates an AI-based credit score for a loan applicant based on financial and demographic data.	Loan Officer, End User	The applicant must have submitted the required financial and personal details.	1. Loan officer enters applicant data. 2. System processes the data. 3. System generates the credit score.	Credit score is generated and displayed to the loan officer.
Loan Eligibility Assessment	The system evaluates whether an applicant is eligible for a loan based on the AI-generated credit score.	Loan Officer	The applicant's credit score must already be generated.	1. Loan officer selects applicant. 2. System analyzes the credit score. 3. System displays eligibility status.	Loan eligibility status is displayed for the applicant.
Fraud Detection	The system flags anomalies in applicant data to detect potential fraud.	System Administrator, Compliance Officer	Applicant data is entered into the system.	1. System monitors applicant data. 2. Anomalies are flagged. 3. Compliance officer is alerted if fraud is detected.	Alerts are sent, and flagged cases are highlighted for review.

Dashboard Reporting	Provides interactive dashboards for loan officers to view credit histories, approvals, and fraud cases.	Loan Officer, Business Analyst	System must have historical data of applicants.	1. Loan officer logs in. 2. System displays dashboard. 3. Loan officer navigates and retrieves required information.	Dashboard data is retrieved and displayed interactively.
System Monitoring	Administrators monitor the performance of the system and ensure it remains functional and secure.	System Administrator	The system is live and processing transactions.	1. Administrator logs into the system. 2. System displays performance metrics. 3. Administrator takes necessary actions if issues are identified.	System is stable, secure, and operational.

Chapter 4

User Interface Specification

Screen Name	UI Elements	Description	Purpose
Signup Screen	Bank logo, Text fields (Full name, Username, Password, Job ID), Dropdown (Department), Submit Button	Allows new users to create an account by entering required details. Department determines whether the user is a customer, loan officer, or admin.	To onboard users into the system and store their credentials securely.
Login Screen	Bank logo, Text fields (Email/Username, Password), Login Button, Forgot Password Link, Create Account Link	Enables existing users to log into the system by entering their email and password.	To provide authorized access to the system.
Entry Details Screen	Bank logo, Text fields (Previous loans, financial data, Personal details, other factors), Save Button, Generate Button	Collects financial and personal data from users, such as income, credit history, employment status, and additional supporting documents.	To gather user information required for credit scoring evaluation.
Dashboard Screen	Bank logo, Navigation Menu, Summary Widgets (Credit Score, Loan Status, Recommendations, Statistics), Notifications, View Details Button	Displays a summary of the user's credit score, loan status, personalized recommendations, and any pending actions or alerts.	To provide users with a quick overview of their credit profile and system recommendations.
Report Screen	Bank logo, Download Button, Print Button, Save Button, Detailed report screen	Offers detailed insights into the user's credit score, loan evaluation, comments from loan officers, and regulatory compliance details.	To allow users download, print, or save the report.

Chapter 5

Domain Analysis

Documentation of the machine learning models, algorithms, and methodologies used in the credit scoring process, including feature engineering, model architecture, and evaluation metrics, supported by diagrams and mathematical representations where applicable.

5.1 Problem Domain

Traditional credit scoring methods are inefficient and prone to bias, relying on outdated manual processes and formulaic approaches that fail to capture the complexity of financial data. These systems require substantial human effort, resulting in delays and errors, while often excluding underserved populations due to a lack of comprehensive evaluation criteria. The proposed AI-based credit scoring system addresses these issues by automating the credit evaluation process, utilizing machine learning to enhance scalability, reduce bias, and ensure more accurate, data-driven decision-making. This innovation seeks to streamline operations for financial institutions and improve access to credit for underserved populations.

5.2 Machine learning specifications

Below are some of the formulas and hyper-parameters as well as a diagram explaining the process flow of a machine learning model.

5.2.1 Hyperparameters

These vary model by model and have to be adjusted to get optimal results via trial and error, following are for a few models to act as a template for what our final model's metrics will look like.

XGBoost/LightGBM:

- learning_rate: Controls step size (e.g., 0.01–0.3).

- n_estimators: Number of boosting rounds (e.g., 100–1000).

- max_depth: Depth of the trees (e.g., 3–10).

Neural Networks:

- hidden_layers: Number and size of hidden layers (e.g., [64, 32]).

- activation_function: Activation functions like ReLU, Sigmoid.

- learning_rate: Optimization step size (e.g., 0.001–0.01).

- batch_size: Mini-batch size (e.g., 32–128).

- epochs: Number of training cycles (e.g., 10–100).

Preprocessing

- scaling_method: Standardization (mean=0, std=1) or MinMax scaling (0–1).

- missing_values_strategy: Mean imputation or predictive imputation.

5.2.2 Evaluation Metrics

- Accuracy: $\frac{TruePositives + TrueNegatives}{TotalInstances}$

- Precision, Recall, F1-Score: For imbalanced data.

- AUC-ROC Curve: To evaluate model discrimination.

- Mean Absolute Error (MAE) / Mean Squared Error (MSE): For regression-based scores.

5.2.3 Formulas

The ML models that we intend to use purely derive decisions based on large amounts of data unlike the following formulas that only take a single record into consideration. These formulas can however be used to assess the performance of ML model against traditional formula-based approach.
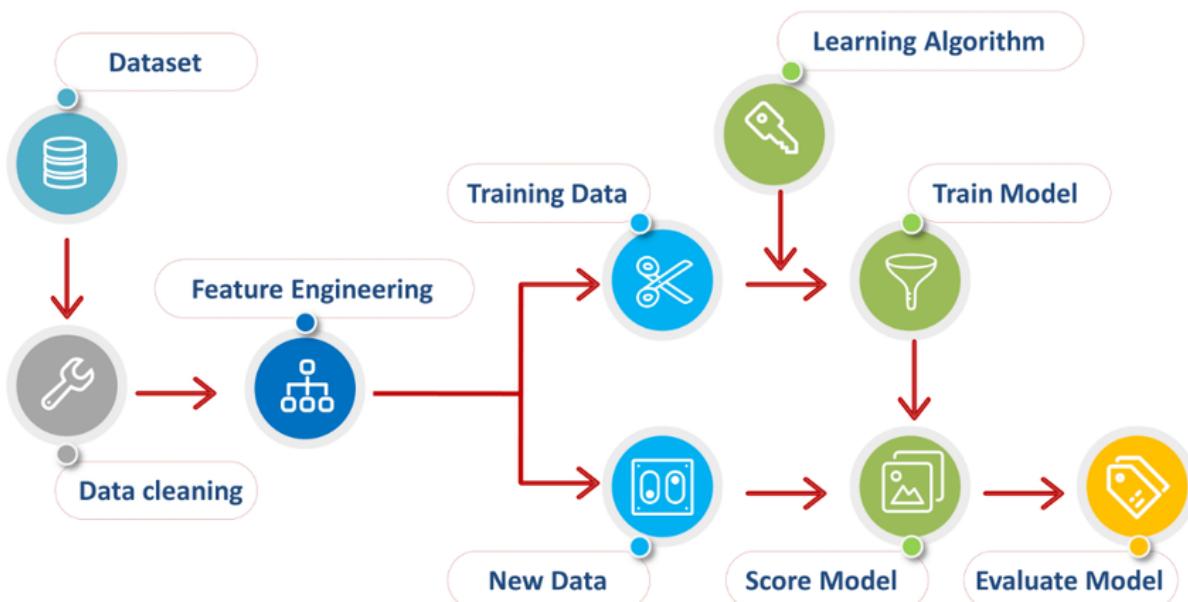
Logistic Regression Formula

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}}$$

Traditional Credit Scoring Formula

$$\text{Credit Score} = \alpha_1 \cdot \text{Income Stability} + \alpha_2 \cdot \text{Debt-to-Income Ratio} + \alpha_3 \cdot \text{Repayment History} + \ldots$$

Machine Learning process

5.3 Stakeholders

The primary stakeholders of the system are financial institutions, such as banks and microfinance organizations, which benefit from accurate and efficient credit evaluations. Loan applicants, particularly those with limited credit histories, gain access to a more inclusive and fair system. Secondary stakeholders include technology providers who enable system implementation, data vendors supplying necessary inputs, and auditors ensuring system integrity and regulatory adherence.

5.4 Data Domain

The system relies on diverse datasets, including transaction histories, financial records, demographic details, and alternative data like social media activity. These inputs are critical for training and operating the AI models. However, challenges such as incomplete or inconsistent data, outliers, and compliance with data privacy regulations need to be addressed. Meezan Bank, the project's partner, will provide historical client data, which must be cleaned, normalized, and securely stored. Real-time data integration further ensures the dynamic adaptability of the system in live environments.

5.5 System Functionalities

The system offers several key functionalities, including automated credit score calculations based on advanced AI models. It provides real-time monitoring and dynamic risk assessment, enabling financial institutions to respond to changes in client behavior swiftly. Fraud detection capabilities are integrated through anomaly detection techniques. Additionally, the platform features customizable dashboards and detailed reporting tools, offering stakeholders actionable insights and regulatory-compliant documentation. These features aim to improve the accuracy, efficiency, and transparency of the credit evaluation process.

5.6 Market and Use Case Analysis

The growing fintech and digital banking sectors in Pakistan underscore the need for innovative credit scoring solutions. The AI-based system addresses key market demands, including creditworthiness assessment for unbanked populations, fraud detection, and real-time credit monitoring. It also enables compliance with Islamic finance principles and regulatory standards. By meeting these needs, the system enhances financial inclusion, reduces risks, and supports the modernization of the credit evaluation landscape in Pakistan.

5.7 Constraints and Assumptions

The project assumes access to sufficient historical and real-time data for training and evaluation. Cooperation from financial institutions is critical for successful implementation and integration into existing systems. However, the system must overcome constraints such as data privacy concerns, scalability challenges, and navigating diverse regulatory environments. Addressing these challenges requires robust data security protocols, flexible system architecture, and proactive stakeholder collaboration.

5.8 Tech Stack

The system employs Python and Flask/Django for backend development, paired with React or Angular for creating interactive dashboards. Data storage is managed through PostgreSQL or MongoDB, ensuring secure and scalable handling of financial records. AI models are developed using TensorFlow and PyTorch. Deployment is facilitated using Docker for containerization. Continuous integration and deployment are supported by GitHub Actions, while data security is enforced through encryption and multi-factor authentication.

Chapter 6

Conclusion

In conclusion, this Software Requirements Specification (SRS) document provides a comprehensive overview of the INTELLISCORE AI-based credit scoring system, detailing its features, user interactions, and technical specifications. The purpose of this document is to ensure a shared understanding of the system's objectives, functionality, and scope among all stakeholders, laying a clear foundation for its development and implementation.

By integrating machine learning and AI algorithms, the system aims to enhance creditworthiness assessments, improve financial inclusion, and reduce operational costs. The SRS also outlines the system's adherence to regulatory standards, ensuring transparency, fairness, and customer satisfaction. Ultimately, this document serves as a guide for building a reliable, scalable, and efficient credit scoring system that will help transform lending practices and broaden access to financial services.