CAPSTONE PROJECT 1

Planning Document

AI-Powered Automation of Credit Risk Report Extraction for MSMEs

by

Ooi Shi Qi

21098272

Bachelor of Information Systems (Data Analytics)

Supervisor: Assoc. Prof. Dr Lee Yun Li

Semester: April 2025

Date: 8 August, 2025

School of Computing and Artificial Intelligence

Faculty of Engineering and Technology

Sunway University

Table of Contents

1.0 Introduction

1.1Background

Micro, Small, and Medium Enterprises (MSMEs) form the backbone of Malaysia's economy, representing a vital driver of employment, growth, and domestic production. These enterprises constitute most firms in the country and significantly form the national business environment. As of 2023, the most recent year with available data, MSMEs accounted for 96.9% of total registered business establishments, reaching 1,101,725 firms nationwide (National Entrepreneur and SME Development Council [NESDC], 2025-a). This widespread presence underscores their centrality in both scale and economic impact.

According to the Department of Statistics Malaysia (2025), MSMEs contributed 39.5% to Malaysia's Gross Domestic Product (GDP) in 2024, with a total value-added of RM652.4 billion at constant 2015 prices, marking a growth of 5.8% from the previous year. This outpaced both national GDP growth (5.1%) and that of non-MSMEs (4.7%), highlighting their dynamism even in a competitive economic environment. MSMEs significantly contributed to sectoral output, accounting for 43.1% of the services sector GDP, 34.1% in manufacturing, and 48.8% in construction, indicating their widespread integration across Malaysia's key industries.

The sector also plays a critical role in employment, accounting for 48.7% of Malaysia's total workforce, or roughly 8.1 million individuals. MSMEs are increasingly engaged in global value chains. In 2024, they contributed 14.3% to total national exports, amounting to RM196.8 billion. It is a notable 31.3% jump from 2023. This surge was driven primarily by the services sector, which alone saw a 114.8% increase in export value (Department of Statistics Malaysia, 2025).

These figures signal the central role of MSMEs in Malaysia's socio-economic development, and by extension, highlight the urgent need to build systems that can support their sustainability and access to essential resources, particularly in areas such as financing, digitalisation, and credit evaluation. Strengthening these pillars ensures that MSMEs remain resilient, competitive, and sustainable in the long term within Malaysia's evolving economic landscape.

1.2 Definition of MSMEs in Malaysia

In Malaysia, MSMEs are officially defined by SME Corp. Malaysia based on two key criteria: annual sales turnover and number of full-time employees. These thresholds vary by sector, distinguishing enterprises across manufacturing, services, and other industries. This classification is structured as shown in Table 1.1.

Table 1.1 Classification of MSMEs in Malaysia

Category	Sector	Sales Turnover	No. of Employees
Micro	All sectors	Less than RM300,000	Fewer than 5
Small	Manufacturing	RM300,000 – RM15 million	5 to <75
	Services & Others	RM300,000 – RM3 million	5 to <30
Medium	Manufacturing	RM15 million – RM50 million	75 to ≤200
	Services & Others	RM3 million – RM20 million	30 to ≤75

Reflecting this classification, MSMEs in Malaysia are predominantly composed of microenterprises, which represented 69.7% of total MSMEs in 2023, amounting to 767,421 firms. These typically include informal businesses such as small roadside food vendors, home-based tailoring operations or family-owned ventures. Small enterprises accounted for 28.5% (314,465 firms), are more commonly found in retail, logistics, and light manufacturing. Medium enterprises, which account for just 1.8% (19,839 firms), typically operate in more formal and capital-intensive sectors such as export manufacturing, IT services, and food processing (NESDC, 2025-b).

This wide spectrum of business sizes and operations reflects the structural diversity within Malaysia's MSME landscape. However, such diversity also implies significant disparities in terms of financial capacity, digital readiness, and formal documentation practices, especially between informal micro firms and more structured medium-sized enterprises (Krishnan & Osman Rani, 2024).

Additionally, MSMEs are geographically concentrated, with Selangor (24.5%), Kuala Lumpur (12.3%), Johor (11.3%), and Penang (7.5%) together accounting for over 55% of total MSME

establishments (NESDC, 2025-b). This geographic clustering reflects stronger access to infrastructure and markets but may also widen the developmental gap between urban and rural-based enterprises.

As a whole, while micro and small firms form the overwhelming majority of Malaysia's business ecosystem, they often operate with limited access to financing and institutional support. This not only constrains their growth but also raises systemic challenges in areas such as formal credit evaluation, where inconsistent documentation and resource limitations hinder accurate and timely decision-making.

1.3 Problem Statement and Motivation

MSMEs are vital contributors to Malaysia's economy but often encounter significant obstacles in accessing formal credit. One of the key obstacles lies in the inefficiencies and inconsistencies of the credit evaluation process, which continues to rely heavily on manual review and interpretation of credit risk documents (Shreya & Pathak, 2025).

1.3.1 Manual credit risk evaluation: a bottleneck

Credit risk reports are essential documents that consolidate a business's financial obligations, repayment behaviour, arrears history, and exposure across multiple lenders. While semi-structured in format, these reports still require credit officers to manually extract and interpret key financial indicators, such as delinquency trends, utilisation limits, and payment histories across multiple pages and formats.

While human judgment remains essential in the credit evaluation process, the upstream task of locating and scrutinizing key information from scanned credit risk reports is often highly manual, repetitive, and time-consuming (Shreya & Pathak, 2025). This dependence on manual data extraction introduces inefficiencies and risks, particularly in institutions facing limited manpower and high volumes of SME loan applications.

Reliance on manual interventions and exception handling can reduce operational consistency and increase the likelihood of process-level errors, particularly under high workload conditions. These challenges point to the need for supportive tools that can streamline data extraction and

presentation, allowing credit officers to focus on validation and decision-making rather than document parsing.

1.3.2 Real-world consequences of inaccurate credit information

Although specific public cases of SME loan rejections caused by manual misinterpretation remain limited, broader industry evidence reveals the consequences of inaccurate credit data in Malaysia. In a notable example reported by Malay Mail, a resort owner had her car loan rejected due to a credit report that incorrectly reflected an outstanding debt. The High Court later ruled in her favour, confirming that CTOS had generated a negative credit record based on inaccurate data, resulting in personal and business losses and a compensation award of RM200,000 (Sharil Abdul Rahman, 2024).

While the case involved personal financing, it highlights the serious downstream impact of incorrect or unverified information in credit reports, whether due to human error, system limitations, or delayed data updates. These risks are magnified when manual workflows dominate credit assessments.

1.3.3 Lack of tools tailored to Malaysian credit reporting formats

Despite the importance of credit risk reports in the evaluation process, most existing automation tools and research are developed for generic or non-Malaysian document formats. These tools often struggle to accommodate the specific layout, terminology, and financial indicators commonly found in Malaysian MSME credit risk reports, which vary significantly in structure across institutions.

Prior studies in other domains have shown that scanned documents with semi-structured layouts, such as medical or laboratory reports, presenting substantial challenges for automated information extraction. Variations in column structure, line positioning, and embedded numeric content often reduce model accuracy if not explicitly handled (Ma et al., 2023; Hsu et al., 2022). These layout-related difficulties are similarly observed in Malaysian credit reports, yet few existing tools have been designed or adapted to parse such formats accurately and efficiently at scale.

1.3.4 Limited usability of existing automation outputs

Even when information extraction is partially automated, many tools produce rigid or non-editable outputs that are disconnected from real-world decision-making needs. Credit officers often find themselves re-validating extracted figures or manually compiling summaries due to unclear formats or missing contextual data, especially for complex indicators such as multi-lender exposure or credit limits.

Practical industry sources confirm that documentation quality remains a major barrier to credit approval. CIMB Business Insights highlights that "poor documentation of business records" is one of the top reasons SME loan applications are delayed or rejected in Malaysia. Similarly, Credit Bureau Malaysia identifies common report inaccuracies such as duplicate accounts, outdated balances, and late reporting of payments, all of which can negatively influence creditworthiness perception (CIMB, 2022; Credit Bureau Malaysia, n.d.-a). These limitations demonstrate that automation alone is insufficient unless paired with intuitive, editable interfaces that support real-world workflows, which is a gap that this project aims to address.

Taken together, these limitations reinforce the need for an AI-assisted, locally tailored solution that can automate credit information extraction. Such a system would reduce the burden of manual review, minimise inconsistency and help ensure more timely, transparent and data-driven credit decisions.

1.4 Project Aim and Objectives

In response to the challenges identified in the previous section, namely the inefficiencies of manual data extraction, limited localisation of existing tools, and the poor usability of automated outputs, this project proposes an AI-assisted system that supports and enhances the review of MSME credit risk reports.

1.4.1 Project Aim

The aim of this project is to develop a system that automates the extraction and structuring of key financial indicators from standardized Malaysian credit risk reports. The system leverages rule-based methods for document-level information extraction and incorporates a Random Forest model to prioritise which financial indicators are most relevant for extraction, based on their significance in MSME credit assessment contexts. The extracted results are delivered

through an interface that supports editing and validation, ensuring that the credit officers remain in control of final decisions.

1.4.2 Project Objectives

To achieve this aim, the project is guided by the following objectives:

1. To identify key financial indicators commonly used in Malaysian MSME credit risk assessment, based on a review of academic literature from the past five years and insights from one domain expert consultation.
2. To implement a Random Forest model for feature prioritisation based on their predictive relevance, thereby informing which indicators should be targeted during the automated extraction process.
3. To design a user-friendly interface that allows users to upload scanned credit risk reports, view and edit extracted data, and export structured outputs for further analysis or internal processing.

This project does not seek to replace human expertise in credit decision-making. Rather, it aims to reduce the burden of manual document review, minimise data inconsistencies, and improve the accessibility of key financial information for use in practical MSME credit evaluation.

1.5 Project Scope and Contribution

1.5.1 Scope

This project is focused on the development of a functional prototype system that automates the extraction of financial indicators from MSME credit risk reports. The scope is deliberately narrowed down to ensure practicality and relevance within the capstone project timeframe.

The system is designed to process scanned PDF versions of standardised credit risk reports issued by Malaysian credit reporting agency, Credit Bureau Malaysia. These documents are typically semi-structured, containing tabular and narrative sections. The extraction logic is implemented using rule-based techniques, but the selection of indicators to extract is guided by a Random Forest model, which ranks financial indicators based on their relevance to MSME credit evaluation.

This project does not aim to develop a full credit scoring system or predictive model. Instead, it focuses on automating the identification and structuring of key financial indicators that are commonly reviewed during credit assessments. The output is presented in a structured, editable dashboard interface that allows users, such as credit officers, to validate, adjust, and export the extracted data for further use.

The system was tested using real-world sample reports made available for academic purposes. These documents reflect the typical structure and content found in Malaysian MSME credit evaluations, allowing for realistic testing of the system's extraction capabilities.

This project is scheduled to last approximately 26 weeks in total. The first 14 weeks will be dedicated to the initiation and planning phases, establishing the framework for the project. The remaining 12 weeks will focus on the execution and closure phases, during which the project plans will be implemented and finalized.

1.5.2 Expected Contribution

This project contributes meaningfully to both the financial technology landscape and the broader MSME credit ecosystem in Malaysia by addressing previously overlooked pain points in credit report analysis and decision support.

Primarily, it delivers a localised digital solution suited to Malaysian MSME credit risk reporting. These reports are commonly used by financial institutions during lending decisions to assess a company's credit exposure, repayment behaviour, and overall financial health. However, they are often overlooked in global AI-based document processing research. This is largely because such reports are proprietary, confidential, and not publicly available. This makes them difficult for researchers to access and study. As a result, most existing systems are trained on generic document types or datasets from Western contexts, such as invoices, receipts, or financial disclosures. By adapting this system to the structure, layout, and terminology of Malaysian credit risk reports, the project fills a practical and persistent gap in current automation tools, many of which perform poorly when applied to local financial documentation.

Second, the system adopts a hybrid AI approach that combines automation with transparency and user control. By using a Random Forest model to prioritise financial indicators, the system ensures that only the most relevant data is targeted during extraction. This helps focus attention on the key variables that truly matter in MSME credit evaluation. At the same time, the use of

rule-based logic for extraction makes the process more predictable and explainable, allowing credit officers to understand how data is being retrieved. This balanced approach not only reduces the time spent manually scanning lengthy reports but also improves the consistency and reliability of the review process.

Third, the system improves practical usability by offering a structured, editable interface where users can review, correct, and export the extracted financial indicators. This solves a common problem in existing tools, which often produce static outputs that cannot be easily adjusted. It is not designed to replace human expertise, but to assist and streamline the review process. Like how spell-check tools aid in document editing without replacing the writer, the system automates repetitive data extraction while leaving the final judgment to credit officers.

Beyond technical contributions, the project has broader economic relevance. As of 2023, MSMEs represented 96.9% of business establishments in Malaysia and contributed 39.1% to the national GDP (NESDC, 2025-b). However, access to formal credit remains a persistent challenge. It is partly due to manual, inconsistent evaluation processes. By improving the efficiency and accuracy of credit assessments, this system supports faster, fairer, and more data-driven financing decisions, ultimately benefiting both financial institutions and MSMEs.

In the long term, such tools can contribute to improving credit access for viable MSMEs, reducing financing bottlenecks, and supporting national goals for economic resilience, digital transformation, and financial inclusion.

2.0 Literature Review

This chapter examines previous research and the theoretical foundations that support this investigation. It begins by reviewing standard credit evaluation processes and financial indicators for MSMEs, then proceeds to discuss machine learning algorithms for feature prioritization. The chapter also analyses the structure and content of Malaysian MSME credit risk reports, along with the shortcomings of current automation approaches. Together, these findings provide the context and rationale for the proposed methodology in this project.

2.1 Credit Evaluation for MSMEs

Credit evaluation is the process of appraising a borrower's risk and creditworthiness. It is a crucial step in deciding if a business qualifies for financing. Traditional credit evaluation frameworks have historically put a high value on audited financial statements, structured financial records, and past repayment patterns.

According to Rajamani et al. (2022), access to financing remains one of the toughest challenges faced by Micro, Small, and Medium Enterprises (MSMEs), especially in developing economies. MSMEs are repeatedly get left out of formal credit channels (Rajamani et al., 2022), as they are unable to satisfy the documentation, collateral or track record criteria imposed by traditional financial institutions. For example, MSMEs are often lack the capacity to compile standardized financial reports or offer strong, sufficient collateral, making them classified as high-risk or unattractive to lenders.

Although these strategies can be effective for large or formally registered enterprises, they often fail to capture the realities faced by MSMEs, whose financial data is typically informal, incomplete, or presented in unstructured formats like scanned reports. As a result, many MSMEs resort to unofficial financing sources, which may meet immediate needs but can limit their potential for long-term growth (Rajamani et al., 2022).

In Malaysian, this issue is extremely acute. Yuan (2020) observes that a large proportion of MSMEs operate informally, rely on cash transactions, and lack audited accounts. These factors create significant information asymmetries, prohibiting viable businesses from accessing credit. As a result, there is growing recognition that traditional approaches must evolve to adapt the

specific documentation patterns and operational realities of MSMEs. The following subsections look at the evolution, from conventional rule-based systems to modern data-driven methods. These set the foundation for understanding how document-level financial data extraction might improve credit evaluation.

2.1.1 Traditional Credit scoring models

Rule-based credit scoring systems are among the oldest and most used methods by financial institutions to assess borrower eligibility. Osman and Rahman (2024) explain that these systems rely on predefined expert logic, often expressed as clear if–then rules. These rules translate lending criteria into structured decision-making frameworks. Their simplicity and transparency make them popular, especially in regulatory settings where consistency and auditability are important.

A well-known example is the 5Cs model. Character, Capacity, Capital, Collateral, and Conditions, which offers a straightforward way to evaluate creditworthiness. This framework is widely used in traditional banking (Osman & Rahman, 2024). Each "C" addresses a specific aspect: Character looks at the borrower's credit history and reliability; Capacity examines income and cash flow to gauge repayment ability; Capital assesses the borrower's net worth or equity as a buffer against risk; Collateral covers assets pledged to secure the loan, reducing lender exposure; and Conditions refer to external factors like economic trends or industry risks, such as inflation or interest rates (Al-Slehat et al., 2024).

While the 5Cs provide a clear, transparent evaluation method, critics argue that they are rigid and may exclude many borrowers. This is especially true for SMEs, which often do not have formal, audited financial records. Osman and Rahman (2024) note that rule-based systems heavily depend on fixed thresholds and historical financial data. When businesses lack such documentation, credit officers must rely more on subjective judgment. This can lead to inconsistencies and biases, particularly when evaluating borderline cases or incomplete information. As a result, some potentially creditworthy borrowers may be unfairly rejected simply because they do not fit the traditional criteria on paper.

These limitations become more pronounced when assessing MSMEs, which are diverse and often operate outside formal financial frameworks. In Malaysia, traditional credit models are still widely used but frequently fail to address the realities faced by micro and small enterprises.

Krishnan and Osman Rani (2024), along with Jalil (2021), highlight that many MSMEs operate informally, conduct business in cash, and lack standardized financial records. Such characteristics conflict with the demands for audited statements, collateral, and formal registration embedded in traditional frameworks. This mismatch creates information gaps and leads to higher loan rejection rates, discouraging MSMEs from seeking formal financing.

Moreover, early-stage and innovation-driven MSMEs often get labelled as high-risk, not because of weak fundamentals but because traditional models overlook alternative indicators of creditworthiness that better reflect their context.

2.1.2 Evolution Towards Data-Driven Credit Evaluation

In response to the drawbacks of traditional credit scoring models, financial institutions are increasingly employing data-driven approaches. The digitization of financial services has made it possible to integrate alternative indicators, such as transactional histories, mobile usage, and behavioural data, particularly for credit assessment in underserved segments (Lomas & Reeta, 2024). This shift has opened the door for machine learning (ML) and artificial intelligence (AI) to create more adaptive, data-rich credit evaluation frameworks.

AI-based models frequently outperform rule-based systems by capturing intricate, non-linear patterns in borrower data, leading to improved prediction accuracy and scalability (Stevenson, 2024). For example, Stevenson (2024) addresses key information asymmetries by showing that transformer-based language models can predict SME default risk from narrative loan documents. Yet despite these advancements, most models remain trained on structured, pre-cleaned datasets.

In real-world scenarios, credit information is frequently embedded in semi-structured formats such as tabular PDFs, scanned reports, or loosely formatted business documents. This creates a practical barrier, as advanced AI models require structured inputs to function effectively. Given the often fragmented and inconsistent nature of MSME financial data, a reliable and interpretable first step is necessary to extract key financial indicators accurately. The next section focuses on the role of extractable financial indicators, with reference to both global and Malaysian contexts.

2.2 Role of Financial Indicators in Credit Risk Assessment

Financial indicators are important in credit risk assessment, as they serve as quantitative proxies for a borrower's repayment ability, financial health and operational stability. These indicators, typically originate from financial statements or credit reports, inform both rule-based scoring systems and machine learning models. Banks, credit bureaus, and fintech companies utilize them extensively to support decisions like interest rate setting, risk grading, and loan approvals. This section reviews commonly adopted financial indicators from both international literatures, as well as Malaysian institutional frameworks.

2.2.1 International Benchmarking and Practices

Across different countries, assessing MSME creditworthiness often combines both financial metrics and non-financial factors. Even though the project's main goal is to automate the extraction of financial indicators, it's crucial to comprehend global evaluation frameworks to determine how the suggested approach fits into standard lending practices.

The use of profitability, liquidity, leverage, and efficiency/activity ratios as key financial indicators in evaluating MSME lending is a recurring theme in the literature. For instance, *Nurani et al.* (2025), studied Indonesian micro-enterprises and identified the debt-to-equity ratio as the primary leverage metric, the current ratio and cash flow for liquidity, and net profit margin, operating margin, and revenue growth for profitability.

In a similar vein, Saygılı et al. (2019), who studied Turkish SMEs, highlighted cash flow to total assets, net income, and the liquidity ratio as key components of financial risk assessment. Kritizopoulos (2019) proposed a more comprehensive set of metrics, focusing on European manufacturing SMEs. These included return on assets (ROA), EBITDA margin, profit per employee, and interest cover. Liquidity was evaluated using the acid-test ratio and cash to current liabilities, while operational efficiency was measured by indicators such as credit period, stock turnover, and collection period.

In the Indian context, *Katoch and Rani* (2023) adopted a longitudinal lens, incorporating credit scores (CIBIL), repayment history, and financial trend analysis to manage non-performing MSME loans. Likewise, *Pambudi* (2022), studying credit underwriting in Indonesian commercial banks, focused on forward-looking indicators such as income patterns, short-term

assets, capital adequacy, and interest coverage ratio, reflecting an emphasis on liquidity and solvency in dynamic market conditions.

A cross-study comparison reveals a consistent subset of financial indicators prioritized in MSME evaluations. Indicators cited in at least three of the five reviewed studies (≥60%) are considered core due to their widespread applicability. The current ratio, mentioned in four studies, stands out as a fundamental liquidity metric. Cash flow, debt-to-equity ratio, and net profit margin appear in three studies each, underscoring their critical role in assessing solvency, leverage, and profitability.

Other indicators such as operating margin, ROA, interest coverage ratio, and revenue growth are cited in two studies, suggesting context-specific but significant value. Less frequently cited indicators, including profit per employee, stock turnover, credit period, and short-term assets, typically reflect sector-specific applications (e.g., manufacturing or commercial banking). While not considered core, these metrics may still hold value in customized credit models.

Table 2.1 summarizes the financial indicators commonly cited across the five international studies.

Table 2.1 Overview of Financial Indicators in Global MSME Credit Assessment Literature

Financial Indicator	**References**				
	(Nurani et al., 2025)	**(Katoch & Rani, 2023)**	**(Pambudi, 2022)**	**(Saygılı et al., 2019)**	**(Kyriazopoulos, 2019)**
Net Profit Margin / Net Income	✓			✓	✓
Operating Margin / EBITDA Margin	✓				✓
Revenue Growth / Income Pattern	✓		✓		
Return on Assets (ROA)					✓

Profit per Employee					✓
Current Ratio	✓		✓	✓	✓
Cash Flow / Cash Surplus or Deficit	✓	✓		✓	
Cash to Current Liabilities / Acid-Test Ratio					✓
Short-term Assets			✓		
Accounts Receivable Ageing		✓			
Expense vs. Income Pattern		✓			
Average Collection Period			✓		✓
Stock Turnover					✓
Credit Period					✓
Debt-to-Equity Ratio	✓		✓		✓
Interest Coverage Ratio / Interest Cover			✓		✓
Capital Adequacy			✓		
Collateral Adequacy / Type		✓	✓		

Credit Score (CIBIL or Bank)		✓	✓		

While financial indicators serve as the primary basis for credit risk assessment, several studies acknowledge the role of non-financial indicators in strengthening lending decisions, especially in contexts where formal financial data may be limited or incomplete. *Nurani et al.* (2025) categorized these into loan, firm, behavioural, management, and market criteria, including factors such as interest rate benefit, bookkeeping behaviour, digital literacy, and participation in fintech ecosystems. Similarly, *Pambudi* (2022) outlined non-financial dimensions such as character (e.g., honesty, credit history), capacity (business revenue and experience), and collateral (asset type and legal clarity).

In India, *Katoch and Rani* (2023) documented rigorous borrower profiling, including Aadhaar-based verification, site visits, and background checks, complemented by behavioural monitoring (e.g., borrower evasiveness, missed communication) and post-disbursement inspections. Operational observations such as irregular site activity and physical stock checks also played a role in risk control.

Table 2.2: Non-Financial Indicators in Global MSME Credit Assessment Literature

Paper	**Non-Financial Indicator Categories**
(Nurani et al., 2025)	• Loan Criteria: interest rate benefit, loan approval rate • Firm Criteria: business size, record-keeping method (manual/digital) • Behavioural: financial discipline, bookkeeping habits • Management: digital literacy, financial training participation • Market: fintech participation, sustainability engagement
(Katoch & Rani, 2023)	• Profile Verification: Aadhaar ID, site visit, background checks • Behavioural: missed communication, evasive borrower behaviour • Operational: irregular site activity, physical stock checks

	• Post-loan Monitoring: phone follow-ups, inspections
(Pambudi, 2022)	• Character: credit history, honesty • Capacity: income, business experience • Collateral: asset liquidity, legal clarity • Condition: market demand, economic trends
(Saygılı et al., 2019)	• Firm Characteristics: firm age, size • Shareholder Profile: age of ownership • Behavioural: export behaviour, operational signals

Although this project does not directly focus on the extraction of non-financial variables, their presence in international MSME credit assessment frameworks underscore the importance of a holistic understanding of creditworthiness. These dimensions such as character, management quality, and behavioural history may inform future system enhancements or hybrid evaluation strategies. A consolidated summary of both financial and non-financial indicators extracted from the reviewed international studies is presented in Table 2.1 and Table 2.2. To contextualize these global practices, the next section examines how credit assessment is conducted within the Malaysian financial ecosystem, highlighting both regulatory frameworks and institutional preferences.

2.2.2 Malaysian Context and Institutional Standards

In Malaysia, the evaluation of creditworthiness for Micro, Small, and Medium Enterprises (MSMEs) incorporates a blend of financial and non-financial indicators, guided by institutional practices, banking standards, and contextual business realities. Several local studies have identified the key indicators that financial institutions typically assess when determining MSME loan eligibility. These findings highlight the balance between regulatory expectations and the pragmatic limitations faced in developing economies, where reliable audited financial statements are hard to come by.

The financial indicators drawn from five Malaysian studies are organized in Table 2.3 according to liquidity, profitability, activity/efficiency, and leverage/solvency.

Table 2.3: Financial Metrics Used in Malaysian MSME Credit Assessment Literature

Category	Indicators	References				
		(Saad et al., 2025)	(Mansor et al., 2023)	(Bakar et al., 2022)	(Abu Hassan, 2020)	(Wasusriyana et al., 2019)
Liquidity	Current Ratio	✓	✓	✓		
	Quick Ratio	✓				
Profitability	Return on Assets (ROA)	✓	✓			
	Profit Margin / Net Profit Margin	✓	✓			
	EBIT			✓		
	Probability Ratios				✓	
Activity / Efficiency	Inventory / Days to Sell	✓		✓		
	Average Collection Period	✓		✓		
	Average Payable Period	✓				
	Cash Flow / Total Asset Ratio	✓				✓
Solvency / Leverage	Debt to Equity Ratio	✓	✓	✓		
	Debt to Asset Ratio		✓	✓		

For liquidity, the current ratio is the most widely cited metric, appearing in three out of five studies (Saad et al., 2025; Mansor et al., 2023; Bakar et al., 2022), as it offers a quick estimate of a firm's short-term financial resilience. Institutional interest in a firm's ability to generate returns relative to its asset base or sales income is reflected in the prominence of profitability indicators such as Return on Assets (ROA) and Net Profit Margin (Saad et al., 2025; Mansor et al., 2023).

According to three studies, leverage metrics, specifically the debt-to-equity ratio—are essential for assessing long-term solvency and financing structure (Saad et al., 2025; Mansor et al., 2023; Bakar et al., 2022). The importance of working capital management and operational efficiency in MSME lending contexts is further highlighted by the frequent appearance of activity-based indicators such as Inventory Turnover and Average Collection Period (Saad et al., 2025; Bakar et al., 2022).

It is also important to note that industry-specific factors can alter the selection and weighting of these ratios. For instance, trading businesses often prioritise turnover and receivables cycles, whereas manufacturing firms are more commonly evaluated in relation to inventory levels and fixed asset utilization (Aulia et al., 2025). Furthermore, some institutions analyse ratio patterns over multiple years rather than relying just on snapshot data, which allows for a more accurate appraisal of financial stability and growth trajectories over time. While the indicators used in Malaysia are locally grounded, they also coordinate with worldwide credit evaluation standards including Basel II/III and IFRS-based SME methods advocated by the World Bank and IFC (Vendy & Sucahyati, 2022; Fišera et al., 2019). This alignment highlights the importance of having reliable and scalable ways to extract these indicators from varied financial documents, especially as Malaysia moves toward greater digital lending and broader financial inclusion.

While the primary focus of this study is on extracting traditional financial indicators, it is important to acknowledge the supporting role of non-financial indicators in institutional decision-making. These qualitative or behavioural elements are often used to complement financial analysis, particularly when financial records are incomplete, outdated, or unaudited.

Table 2.4 presents a summary of the non-financial indicators identified in Malaysian MSME credit evaluation literature, categorized by study. These factors ranging from managerial characteristics to behavioural and market considerations. They highlight the multi-dimensional nature of credit assessment in real-world lending contexts.

Table 2.4: Non-Financial Indicators in Malaysian MSME Credit Evaluation

Paper	Non-Financial Indicators Reported
(Saad et al., 2025)	• Loan-related: existing bank-firm relationship, collateral availability • Firm Characteristics: business age, legal status, ownership of premises • Management: owner's education and experience • Behavioural: tax compliance, past payment history • Market: technology adoption, competitive positioning
(Mansor et al., 2023)	• Credit Profile: institutional credit rating • Operational Flexibility: ability to adapt production to demand shifts
(Bakar et al., 2022)	• Management: business owner's experience and education • Firm Demographics: gender, business age
(Abu Hassan, 2020)	• Collateral: type of collateral offered, legal enforceability • Firm Status: registration status, industry classification
(Wasiuzzaman et al., 2019)	• Business Form: sole proprietorship, partnership, etc. • Behavioural: repayment conduct, payment history • Collateral: asset liquidity, encumbrance • Condition: product demand, industry outlook

For instance, **loan-related criteria** such as the existence of a **bank–firm relationship** or the availability of **collateral** remain critical in traditional credit evaluation (Saad et al., 2025; Abu Hassan, 2020). Similarly, **firm-level criteria** such as **business age**, **legal structure**, and **ownership of premises** are considered proxies for operational stability. Studies have also emphasized **managerial attributes**, including the **experience, education, and performance** of business owners or directors, which are perceived as influencing business sustainability (Saad et al., 2025; Bakar et al., 2022).

Behavioural indicators such as **tax compliance**, **past payment history**, and **number of loans or arrears** reflect a firm's historical credit discipline (Wasusriyana et al., 2019). Some institutions also consider **market-oriented factors** like **technology adoption** and **competitive awareness**, particularly when evaluating businesses in innovation-driven or rapidly evolving sectors (Saad et al., 2025, Wasusriyana et al., 2019).

Although some of these non-financial indicators are not targeted in the present study's extraction system, they provide valuable context and may inform future extensions of the pipeline, especially for hybrid scoring models or explainability-enhanced credit analysis.

In summary, the Malaysian institutional landscape for MSME credit assessment reveals a strong alignment with globally accepted financial metrics, supplemented by contextual non-financial considerations. This dual-layered approach underscores the importance of designing AI-driven extraction systems that prioritize structured financial data while remaining adaptable to the semi-structured and nuanced nature of credit risk reports.

2.2.3 Consolidated Financial Indicators from Literature Review

Drawing from the cross-analysis of international and Malaysian credit assessment literature in sections 2.2.1 and 2.2.2, this section consolidates a refined list of financial and selected non-financial indicators that are both most frequently cited and are feasibly extractable from standardized credit reports. These indicators form the core feature set for downstream analysis and machine learning prioritization in later methodology.

Indicators were shortlisted using two distinct criteria:

1) **Literature Support:** The indicator appears in three or more academic studies, based on the combined count from both global and Malaysian sources. This reflects wide academic consensus and applicability across different credit evaluation contexts.
2) **Report Availability:** The indicator is structurally present in formal Malaysian credit risk documents such as *MyBizScore*, enabling reliable rule-based extraction.

Most financial indicators fall within conventional categories such as liquidity, profitability, leverage, and efficiency. In addition to these, two non-financial indicators, **business age** and **repayment conduct** are also included. These were chosen because they meet both criteria: they are widely cited in the literature and consistently appear in structured formats within credit reports.

Table 2.5 summarizes each indicator's academic support and its presence in *MyBizScore*. Indicators that meet both criteria are prioritized in the current study.

Table 2.5: Summary of Indicator Support and Availability

Indicator	**Backed by Academic Papers**	**Present in MyBizScore Report?**
Current Ratio	✓	✓
Net Profit Margin	✓	✓
Debt-to-Equity Ratio	✓	✓
Return on Assets (ROA)	✓	✓
Average Collection Period	✓	✗
Quick Ratio	✓	✗
Operating Margin / EBITDA	✓	✗
Inventory Turnover	✓	✗
Interest Coverage Ratio	✓	✗
Business Age	✓	✓
Repayment Conduct	✓	✓

Although some indicators, such as Net Profit Margin and Debt-to-Equity Ratio, are not explicitly stated in the credit report, they can be derived from structured base values (e.g., revenue, net profit, liabilities, equity). These are retained in the list as they can be reliably computed during extraction using simple backend logic.

Several other indicators, such as the quick ratio, interest coverage ratio, and inventory turnover, are cited in academic literature but are excluded from the current feature set due to their absence in structured Malaysian credit reports. These may be considered in future work if data quality and availability improve.

Table 2.6: Indicator Descriptions and Relevance

Indicator	Descriptions / Meaning	Why It is Used
Current Ratio	Current Assets ÷ Current Liabilities	Measures liquidity and ability to meet short-term obligations
Net Profit Margin	Net Profit ÷ Revenue	Indicates operational profitability and cost efficiency
Debt-to-Equity Ratio	Total Liabilities ÷ Shareholders' Equity	Evaluates leverage and financial risk
Return on Assets (ROA)	Net Income ÷ Total Assets	Measures efficiency in asset utilization
Business Age	Years since company registration	Proxy for operational maturity and business experience
Repayment Conduct	Month-by-month loan conduct history (12 months)	Captures behavioural reliability and delinquency risk

The following section explores how the shortlisted indicators are ranked using machine learning techniques to identify which variables contribute most to credit evaluation outcomes.

2.3 Feature Prioritization Using Machine Learning

Effective credit risk evaluation often involves analysing a multitude of financial and behavioural indicators. However, not all variables equally contribute to predictive accuracy or operational relevance. Some features may be redundant, weakly correlated, or even introduce noise that degrades model performance. Therefore, prioritizing features by identifying the most predictive indicators is crucial to improve both model interpretability and efficiency in real-world lending applications, particularly for MSMEs. This section reviews techniques used in prior research to rank financial indicators.

2.3.1 Overview of Feature Ranking Techniques in Credit Scoring

Machine learning-based feature ranking methods have gained prominence in credit risk research due to their ability to model complex, nonlinear relationships inherent in financial data. Commonly adopted approaches include:

- Statistical Filter Methods:

  Traditional methods such as Information Gain and Chi-square tests, function as preliminary screening mechanisms that rank features according to their marginal relevance to the target variable. Owing to their model-agnostic and computationally efficient nature, they are well-suited for rapid dimensionality reduction. Nonetheless, their principal limitation lies in the univariate evaluation of features, which overlooks inter-feature dependencies and disregards the structural nuances of the predictive modelling context, thereby potentially undermining downstream model performance (Laborda & Ryoo, 2021; Jemai & Zarrad, 2023).

- Wrapper Methods:

  Recursive Feature Elimination (RFE) are designed to iteratively identify optimal feature subsets that maximize predictive accuracy by removing less informative variables. While RFE is commonly applied with models like SVM or logistic regression in broader literature, Koç et al. (2023) and Laborda & Ryoo (2021) adopt other wrapper approaches, such as sequential forward or stepwise selection. Both similarly capture feature dependencies but tend to be computationally intensive and susceptible to overfitting in small datasets.

- Model-Agnostic Explainability Techniques:

  SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) have two recent examples of effective tools for evaluating black-box models. They assign relevance scores by quantifying each feature's contribution to individual predictions, resulting in for more transparent decision-making in credit risk assessment.

  (Wibowo & Sanjaya, 2024; Shreya & Pathak, 2025).

- Tree-Based Feature Importance:

  Ensemble tree models such as Random Forests (RF), Gradient Boosting Machines (e.g., XGBoost), and Extremely Randomized Trees inherently produce feature importance metrics during training by measuring the frequency and effectiveness of each variable in splitting decision trees to reduce impurity. These models work effective with mixed types and missing values in financial data. (Osman & Rahman, 2024; Jemai & Zarrad, 2023; Babaev et al., 2019).

Table 2.7 Comparison of Feature Ranking Methods

Method Category	Examples	Key Strengths	Main Limitations	Typical Applications
Statistical Filter Methods	Information Gain, Chi-square	Fast, simple, low computation	Ignore interactions, redundancy	Initial screening, dimensionality reduction
Wrapper Methods	Recursive Feature Elimination (RFE)	Considers feature dependencies	Computationally expensive, overfitting risk	Fine-tuning feature subsets for specific models
Model-Agnostic Explainability	SHAP, LIME	Explains black-box models, local/global importance	Computationally heavy, complex for users	Transparency in complex ensemble models
Tree-Based Feature Importance	Random Forest, XGBoost	Built-in importance, handles mixed data, robust	Bias to features with many splits	Feature ranking in tabular financial data

A combination of feature selection techniques, including statistical filters, wrapper-based algorithms, and tree-based importance measures, can improve model robustness and facilitate

dimensionality reduction, as empirically demonstrated by Jemai and Zarrad (2023). This is accomplished by eliminating redundant or unnecessary variables without sacrificing predictive power. Additionally, Shreya (2025) and Wibowo and Sanjaya (2024) highlighted the growing use of SHAP and LIME for explainability in credit scoring, particularly when employing sophisticated ensemble learners, thereby enhancing regulatory compliance and stakeholder confidence.

2.3.2 Advantages of Random Forest for Feature Prioritization in Financial Risk

Due to its numerous practical advantages, Random Forest (RF) has become one of the most widely used machine learning models for feature ranking in credit risk assessment. Without extensive preprocessing, RF can automatically handle datasets containing both numerical and categorical variables. It is also resilient to noisy or missing data, which are common issues in MSME credit datasets. Unlike linear models, RF efficiently addresses multicollinearity among predictors without significantly degrading predictive performance, making it reliable in financial contexts where features are often correlated. Furthermore, RF automatically produces feature importance scores based on information gain or Gini impurity measures, facilitating the ranking and selection of the most predictive variables. These features help make RF appropriate for small to medium-sized datasets, which are common in MSME research, where few samples and class imbalance can otherwise compromise model stability. Even with restricted data, the ensemble nature of RF mitigates overfitting and provides consistent, interpretable variable importance rankings (Breiman, 2001).

Numerous recent studies have demonstrated the practical value of RF in prioritizing credit risk features. Within an ensemble framework, Osman and Rahman (2024) successfully identified critical financial variables predictive of SME loan approvals by combining RF with deep learning models. Their findings showed that leveraging RF's feature ranking capabilities significantly enhanced classification accuracy and interpretability. Similarly, Babaev et al. (2019) highlighted RF's interpretability and robustness in credit scoring tasks, emphasizing its ability to model complex nonlinear interactions without sacrificing transparency. Compared to more conventional models, RF's advantage in ranking key credit risk variables was further validated empirically by Jemai and Zarrad (2023). Moreover, Wibowo and Sanjaya (2024) reinforced RF's crucial role in realistic financial risk processes by combining it with SHAP

explainability techniques to produce clear and interpretable feature rankings for MSME non-performing loan prediction.

2.3.3 Explainable AI Techniques for Enhanced Feature Interpretation

In order to improve the transparency of feature importance evaluations, recent advances in explainable AI (XAI) have paired tree-based learners with model-agnostic techniques such as SHAP and LIME. SHAP provides both global and local interpretability by attributing the contribution of each input feature to individual predictions using game-theoretic principles.

Wang and Liang (2024) conducted a thorough evaluation of interpretability methods in credit scoring, finding that LIME offers better end-user comprehension, while TreeSHAP delivers the most consistent and reliable feature importance estimates. These findings align with those of Shreya (2025), who successfully applied SHAP and LIME to ensemble models like RF and XGBoost, producing valuable insights into key credit risk indicators.

Similarly, Wibowo and Sanjaya (2024) illustrated how boosting models combined with SHAP explanations enable fine-grained understanding of feature effects in MSME loan default prediction. This integration of XAI enhances model transparency and supports regulatory requirements for explainability in financial decision-making.

2.3.4 Summary

Overall, the literature analysis highlights the effectiveness of machine learning methods for feature prioritization in credit risk assessment, with Random Forest emerging as a practical and interpretable baseline approach. Explainable AI tools such as SHAP and LIME have been shown to further enhance feature importance analysis by providing clear yet detailed interpretations of model predictions. However, considering real-world constraints such as time and resource availability, this study focuses on leveraging Random Forest's built-in feature ranking capabilities without incorporating additional explainability techniques. This method provides a strong foundation for identifying key financial indicators within the research context, balancing methodological rigor with practical considerations.

2.4 Structure and Content of MSME Credit Risk Reports

2.4.1 Role and Importance of Credit Risk Reports

Before offering loans or credit facilities, financial institutions rely on credit risk reports as essential tools for assessing borrowers' creditworthiness. These reports provide a comprehensive view of a borrower's financial commitments, repayment history, and current exposure, enabling lenders to estimate default probability and make well-informed lending decisions (Credit Bureau Malaysia, n.d.-c). For many Micro, Small, and Medium-Sized Enterprises (MSMEs) in Malaysia, which often operate with limited or fragmented official financial records, such reports are particularly crucial (Krishnan & Osman Rani, 2024). In these circumstances, standardized credit reports serve as the primary source of structured financial information.

2.4.2 Major Credit Reporting Systems in Malaysia

The Central Credit Reference Information System (CCRIS), developed and administered by Bank Negara Malaysia (BNM), is the most widely recognized credit reporting system in the country. CCRIS compiles credit data from participating financial institutions into a centralized database. In addition to 12-month payment histories, it provides a summary of the borrower's existing loans and credit facilities. According to Bank Negara Malaysia (n.d.), the report contains key details such as loan type, disbursement amount, monthly installment, current balance, and legal enforcement status. It is widely used across the Malaysian financial sector to assess a borrower's exposure across banks and evaluate their repayment ability, particularly during loan application reviews.

In addition to CCRIS, credit reports are often provided in PDF format by private organizations such as Credit Bureau Malaysia (CBM) and CTOS Data Systems. These reports typically combine CCRIS data with third-party information and proprietary algorithms. CBM's MyBizSCoRE is a prominent MSME-focused report tailored for institutional lending, featuring a structured credit scoring system based on Probability of Default (PD), trade reference data, financial summaries, directorship and shareholder details, CCRIS-derived banking information, and legal actions (Credit Bureau Malaysia, n.d.-b).

2.4.3 Typical Report Structure

The typical sections included in a MyBizSCoRE report is presented in **Table 2.8**, which outlines each section's content, format, and relevance to credit risk evaluation.

Table 2.8 Typical Structure of a CBM MSME Credit Risk Report

Section	Contents / Subsections	Format	Credit Risk Relevance
1. Report Summary	SME Credit Score, CCRIS summary, Trade & Litigation alerts, Key risk indicators	Structured summary with visuals	Offers an instant risk snapshot; helps prioritize detailed analysis
2. SME Profile & Corporate Information	Company name, business registration, constitution, address, directors/shareholders	Structured tables/lists	Establishes legal identity, ownership, and governance context
3. Share Capital & Charges	Authorized/issued capital, shareholder equity, secured charge registrations	Tabular data	Highlights capital structure and financial obligations
4. Financial Statements	3–5 years of balance sheets, income statements, financial ratios (e.g., ROCE, gearing ratio, EPS)	Time-series tables	Indicates financial health trends critical for solvency and performance evaluation
5. Banking Information	CCRIS-derived facilities: loan types, lenders, outstanding balances, repayment conduct, dishonoured cheques, facility limits	CCRIS-based tables	Vital for exposing exposure levels, repayment behaviour, and potential defaults
6. Trade Credit Reference	Supplier payment behaviour, aging analysis, trade defaults	Tables with narrative	Offers insight into non-bank credit reliability and payment discipline
7. Litigation / Winding-Up Information	Legal cases, court orders, insolvency or winding-up proceedings	Structured legal text	Indicates reputational and regulatory risks affecting creditworthiness

A visual example of the report is provided in **Appendix A**, which includes annotated screenshots to illustrate how key data elements are organized in practice.

2.4.4 Differences Between CBM and CTOS Reports

While both CBM and CTOS issue business credit reports, their structures and reporting emphases differ. CTOS reports tend to focus on SSM company registry data, visual litigation summaries, and vendor risk insights, often presented with more narrative descriptions and visual flags (CTOS Data Systems, n.d.-b). They are primarily intended to support general business risk assessment and to facilitate due diligence processes when evaluating potential customers, suppliers, or business partners. (CTOS Data Systems, n.d.-a).

In contrast, CBM's MyBizSCoRE report is optimized for institutional credit decisions, offering structured financial metrics and risk scoring formats that support automation, standardization, and integration with machine learning-based assessment methods, aligning closely with the objectives of this study.

2.4.5 Layout Variability and Automation Challenges

The MyBizSCoRE credit risk reports are digitally generated PDFs featuring a well-structured and consistent layout, including clearly labeled sections and organized tabular formats. This structure facilitates systematic data extraction without the need for Optical Character Recognition (OCR), as the content is already machine-readable.

However, additional sections, missing data, or changes between report versions may lead to slight layout variations. Flexible parsing techniques are required for tables containing multi-line cells and fields of varying lengths. Furthermore, bilingual content, primarily limited to addresses and disclaimers, adds minimal complexity, as it typically does not affect key financial elements.

These minor discrepancies and the need to handle incomplete entries while preserving data quality, represent the main automation challenges rather than fundamental structural changes. Sensitivity to these subtleties is essential when designing reliable extraction logic to guarantee accurate and transparent credit evaluations.

2.4.6 Justification for Rule-Based Extraction

It is both feasible and efficient to use a rule-based extraction strategy because of the consistent structure and digital format of Malaysian MSME credit risk report. As opposed to OCR or machine learning methods, which work better to unstructured or image-based documents. Rule-based techniques offer direct parsing to text and rely on identifiable patterns such as fixed headers and table layouts.

This approach aligns with the project's goals of transparency, reliability, and ease of customization. When used on stable formats, rule-based approaches are simpler to maintain and provide explicit logic paths.

Rijcken et al. (2025), who examined rule-based and machine learning approaches for text classification in structured domains, provide more evidence in favor of this conclusion. Their findings indicate that rule-based systems perform competitively, sometimes outperforming ML models, particularly in environments where document structure is consistent and domain knowledge is available. They highlight key advantages such as lower data requirements, easier validation, and greater interpretability.

Therefore, in the context of standardized credit risk reports, rule-based extraction not only reduces system complexity but also ensures auditable and accurate results, aligning with both the document characteristics and the broader objectives of this project.

2.5 Gaps in Existing Literature

2.5.1 Technical Limitations in Existing Automation Research

While there is considerable research on automating credit evaluation, much of it focuses on relatively generic financial documents such as invoices, receipts, or well-structured financial statements (Bhatt, 2022). However, no studies have specifically tackled the challenges involved in extracting information from credit risk reports, especially those issued by Malaysian agencies like Credit Bureau Malaysia (CBM). These reports present a mixture of structured financial tables, semi-structured narratives, and often bilingual content in Malay and English, increasing the complexity of automated extraction. Furthermore, many prior studies assume access to clean, consistently formatted datasets, which is rarely the case with actual Malaysian MSME credit risk reports. Common automated techniques relying on Optical Character Recognition (OCR) and Natural Language Processing (NLP) often face difficulties when handling such complex layouts, mixed languages, and incomplete data. These technical

shortcomings highlight the need for extraction methods customized to the distinctive features of Malaysian credit risk documents, an approach this project seeks to explore.

2.5.2 Practical Challenges in Automated Extraction Outputs

Many current automated extraction tools produce outputs that are inflexible and difficult to adapt, limiting their usability in practical credit assessment contexts. In real-world workflows, credit officers must frequently review, validate, and adjust extracted data to compensate for issues such as incomplete documentation, outdated information, or duplicate records. Without user-friendly interfaces that support such manual interventions, automated outputs risk propagating errors and potentially compromising credit decision accuracy. This challenge is recognized by both financial institutions and credit reporting agencies, underscoring the importance of systems designed to balance automation with human oversight (CIMB, 2022; Credit Bureau Malaysia, n.d.-c).

2.5.3 Contextual Gaps: The Malaysian MSME Landscape

From a contextual perspective, Malaysian MSMEs, particularly microenterprises like home-based businesses and small retailers, often operate with limited formal financial documentation. This is typically due to constraints such as restricted financial resources, lack of trained accounting staff, and limited access to digital accounting tools. As a result, lenders in Malaysia rely extensively on credit risk reports produced by agencies like CBM, which aggregate financial data, repayment histories, and other relevant credit information from multiple sources. This reliance contrasts with developed economies where MSMEs usually maintain formal accounting records that enable lenders to base their assessments on detailed and regularly updated financial statements. Despite the importance of these credit reports, there is a notable gap in automation research addressing the specific structure and content of Malaysian credit risk reports (Krishnan & Osman Rani, 2024; Yuan, 2020).

2.5.4 Limited Practitioner Involvement in Financial Indicator Validation

The limited incorporation of practitioner feedback in validating financial indicators is another significant gap in the literature. Few studies incorporate perspectives from credit officers or other subject matter experts to confirm the practical relevance of these indicators, even though academic research offers a broad range of theoretically significant financial measures. This lack of real-world validation often results in models or extraction systems that do not align

closely with institutional procedures or decision-making requirements. To develop reliable and practical credit evaluation systems, empirical research that integrates literature-based feature identification, expert validation, and data-driven prioritization is urgently needed.

2.5.5 Research Opportunity

When taken as a whole, these contextual and technical limitations highlight the urgent need for targeted, locally relevant research tailored to the Malaysian MSME credit appraisal context. This project aims to address these shortcomings by developing a flexible extraction framework, integrating an intuitive dashboard for expert interaction, and implementing a multi-layered feature prioritization strategy that combines findings from the literature review, expert insights, and artificial intelligence techniques.

2.6 Chapter Summary

The foundation for the suggested AI-assisted system that automates the extraction of financial indicators from Malaysian MSME credit risk reports is established by this review of the literature. It first contextualizes the challenges MSMEs face in credit evaluation, emphasizing limitations inherent in traditional rule-based methods and highlighting the shift towards data-driven strategies leveraging machine learning techniques.

In credit risk assessment, the central role of financial indicators was thoroughly addressed, reviewing both international and Malaysian contexts. Financial indicators such as Current Ratio, Debt-to-Equity Ratio, Return on Assets, Net Profit Margin, Business Age, and Repayment Conduct were consistently highlighted across multiple studies as critical metrics influencing MSME credit evaluation.

Subsequently, methodologies for feature prioritization using machine learning, particularly Random Forest models, were discussed. The review justified Random Forest's suitability based on its robustness, interpretability, and compatibility with MSME-related data characteristics. Despite discussing explainability techniques such as SHAP and LIME, practical constraints dictate that this project will initially rely solely on the inherent interpretability provided by Random Forest for feature selection.

The structure and content of Malaysian credit risk reports were analyzed, clearly justifying the feasibility of a rule-based extraction approach given their standardized yet semi-structured

layout. This review concluded by explicitly identifying technical and contextual gaps in existing literature, notably highlighting the limited applicability of existing automated methods to Malaysian MSME credit documentation, the absence of flexible, editable interfaces, and the rarity of integrating expert validation in feature selection.

These identified gaps directly justify and guide the methodological decisions described in the subsequent chapter, establishing a clear and coherent transition into the proposed extraction and prioritization strategy. Thus, the literature review provides both theoretical validation and practical direction for the AI-powered system this project seeks to develop.

3.0 Methodology

This chapter explains the methodological approach undertaken to build the AI-powered system and tested it to automate the extraction of financial indicators from MSME credit risk reports. The methodology is structured around addressing the core problem of manual data extraction by presenting the entire workflow, the processes of identifying key features, system development phases, and evaluation strategies. Each section discusses the tools, techniques, and decisions involved, ensuring the solution is both technically sound and suitable for real-world credit assessment needs.

3.1 Problem Understanding

Manual extraction of financial indicators from MSME credit risk reports remains a common but inefficient practice in many financial institutions. Loan officers are often required to review large volumes of loan application documents, which makes the process time-consuming, error-prone, and inconsistent. These inefficiencies hinder scalability and introduce delays in credit decision-making, disproportionately affecting MSMEs that typically face greater barriers in accessing timely financing. While digital lending platforms are on the rise, many automation approaches assume access to clean, structured data, an assumption that does not hold true in the case of real-world MSME documentation, particularly in developing economies like Malaysia (Osman & Rahman, 2024). Furthermore, existing tools often lack the flexibility, transparency, and adaptability needed to align with practical credit evaluation workflows (Osman & Rahman, 2024). There is a clear need for lightweight, AI-assisted system capable of accurately extracting key financial indicators from credit reports in a way that is both interpretable and editable by human users. This project responds to that need by developing a semi-automated extraction tool that integrates expert-informed feature design, machine learning-driven feature selection, and user-driven interface validation.

3.2 Project Overview and Workflow

The system addresses current inefficiencies in manual credit evaluation, where credit officers often rely on subjective judgment and time-consuming review of various financial documents, many of which are semi-structured or inconsistently formatted. This project, however, narrows its scope specifically to MSME credit risk reports, which follow a more consistent layout but still require manual data extraction.

A workflow diagram, figure 3.1 is provided to visually represent the overall process from input ingestion to final evaluation.

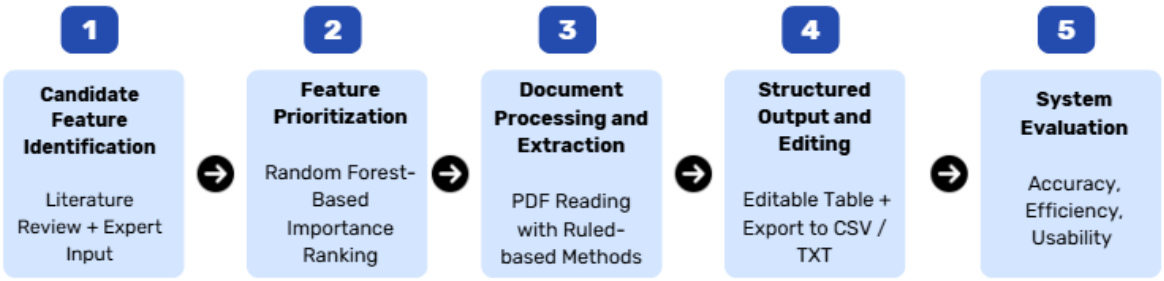


Figure 3.1: Project Workflow Overview

The methodology is structured into five key phases:

Step 1 - A comprehensive set of candidate financial indicators are identified through academic literature (both global and Malaysian contexts) and validated with domain expert input. This ensures that the system retrieves meaningful and contextually relevant variables.

Step 2 - A Random Forest model is trained using labelled historical data to rank the importance of candidate indicators in predicting credit scores. This prioritization guides which fields are most valuable to extract during document processing.

Step 3 - Credit risk reports in PDF format are processed using rule-based techniques such as pdfplumber.extract_tables() and regular expressions. The system anchors to section titles (e.g., "FINANCIAL STATEMENT", "LOAN INFORMATION") to locate and extract only the selected indicators, ensuring consistent and targeted extraction across varying layouts.

Step 4 - A Streamlit dashboard displays the extracted indicators in an editable, structured table. Users can review and modify the values using their subject-matter expertise. The finalized dataset can then be exported in TXT or CSV format for later use.

Step 5 - The system's usability, time efficiency, and extraction accuracy are evaluated. To ensure the solution is effective and applicable for real-world credit assessment, these aspects are assessed through task-based testing and expert surveys. This includes verifying the accuracy of the rule-based extraction process as well as the relevance of the Random Forest feature ranking.

3.3 Identification of Extractable Features

The primary objective of this phase is to identify a comprehensive set of financial indicators relevant to MSME credit risk assessment. The goal is to develop a candidate feature list that combines real-world data from academic research in both Malaysian and international contexts with practical insights from industry professionals. This carefully curated feature set will form the basis for subsequent processes, including rule-based data extraction, feature prioritization using AI techniques, and system performance evaluation.

3.3.1 Feature Collection from Literature Review

To develop a foundational list of financial indicators often used in credit scoring and loan eligibility assessment, an extensive literature review was conducted. From reputable journals and conferences, scholarly articles focused on research on MSME financing, credit evaluation models, and AI-based financial risk systems were selected. The primary inclusion criterion for an indicator to be shortlisted was that it must be cited in two or more independent sources, to ensure empirical support and relevance. Indicators that were specifically applied to MSMEs or aligned with Malaysia's credit assessment context were prioritized, although international sources were also considered to widen the scope.

To enhance contextual clarity, the identified indicators were grouped into two categories:

(i) those drawn from Malaysian studies, and
(ii) those derived from global studies.

This grouping allows for a clearer comparison of regional versus international practices and will support later interpretation of whether certain indicators are more locally grounded or globally standardized. A final merged comparison table, consolidating both streams, is presented below in Table 3.1.

Table 3.1 Merged Comparison of Global and Malaysian MSME Credit Assessment Indicators

	Global Studies	**Malaysian Studies**
Financial Indicators	• Net Profit Margin / Net Income	• Current Ratio • Quick Ratio

	• Operating Margin / EBITDA Margin • Revenue Growth / Income Pattern • Return on Assets (ROA) • Profit per Employee • Current Ratio • Cash Flow / Cash Surplus or Deficit • Cash to Current Liabilities / Acid-Test Ratio • Short-term Assets • Accounts Receivable Ageing • Expense vs. Income Pattern • Average Collection Period • Stock Turnover • Credit Period • Debt-to-Equity Ratio • Interest Coverage Ratio / Interest Cover • Capital Adequacy • Collateral Adequacy / Type • Credit Score (CIBIL or Bank)	• Return on Assets (ROA) • Profit Margin / Net Profit Margin • EBIT • Probability Ratios • Inventory / Days to Sell • Average Collection Period • Average Payable Period • Cash Flow / Total Asset Ratio • Debt-to-Equity Ratio • Debt-to-Asset Ratio
Non-Financial Indicators	• Loan Criteria (e.g., interest rate benefit, loan approval rate)	• Loan-related (e.g., existing bank relationship, collateral availability)

	• Firm Characteristics (e.g., business size, record-keeping method) • Behavioural (e.g., financial discipline, export behaviour) • Management (e.g., digital literacy, training participation) • Market (e.g., fintech participation, market demand) • Profile Verification (e.g., site visit, background checks) • Operational (e.g., irregular site activity) • Capacity (income, business experience) • Collateral clarity • Shareholder profile	• Firm Characteristics (e.g., business age, legal status) • Management (e.g., owner's experience, education) • Behavioural (e.g., tax compliance, repayment conduct) • Market (e.g., technology adoption, competitive positioning • Credit Profile (institutional rating) • Operational Flexibility • Firm Demographics (e.g., gender) • Collateral type and enforceability • Firm Status (e.g., registration status) • Business Form (e.g., sole proprietorship, partnership)

3.3.2 Feature Collection from Expert Input

To complement the literature-derived feature list and ensure alignment with real-world practices, domain expert validation was conducted using a structured Google Form survey. The objective was to identify which financial and non-financial indicators are regularly considered by practitioners when evaluating the creditworthiness of MSMEs. The form included multiple-choice sections structured by typical credit risk report categories and an open-ended section for

additional suggestions. The following analysis interprets the expert's selections across each section to highlight validated indicators and identify any additional practitioner-driven insights.

Expert Profile

One expert participated in this initial round of validation to assess the practical relevance of financial indicators used in MSME credit evaluation. The respondent, Mr. Ooi Eng Huat, currently serves as the Company Director of OEH Consulting Services (M) Sdn Bhd. With between six and ten years of experience in credit evaluation, Mr. Ooi brings solid mid-to-senior level expertise shaped by hands-on involvement in financial consulting for MSMEs. His industry focus falls within the consulting and advisory services sector, a field that often requires tailored assessments and close interaction with MSME clients. This context is significant as it reflects a practitioner's viewpoint rooted not only in theoretical knowledge but also in day-to-day creditworthiness assessments. Prior to participation, informed consent was obtained, permitting the use of his responses for academic research purposes.

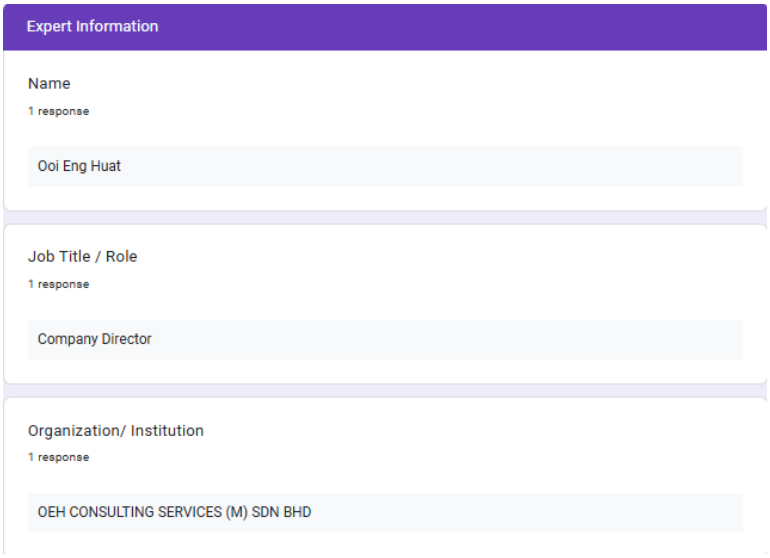


Figure 3.2: Expert Respondent's Professional Information

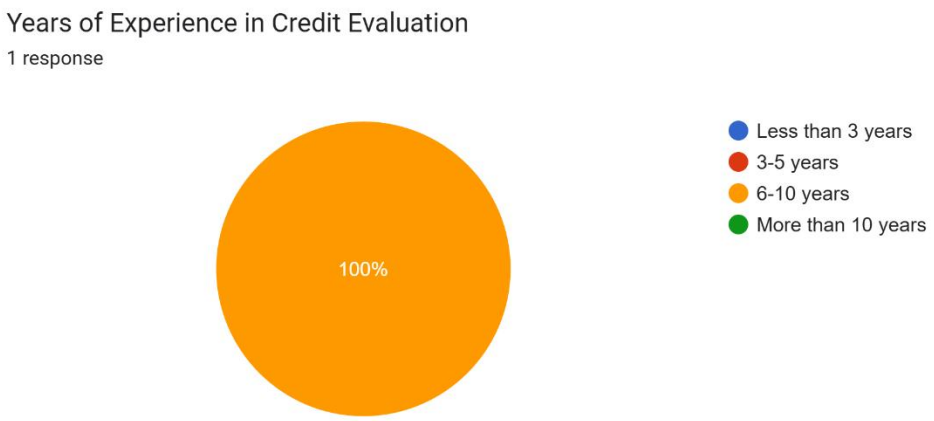


Figure 3.3: Years of Experience in Credit Evaluation

As shown in Figure 3.3, the respondent selected the "6–10 years" category, confirming a solid mid-level to senior experience range in credit evaluation practices.

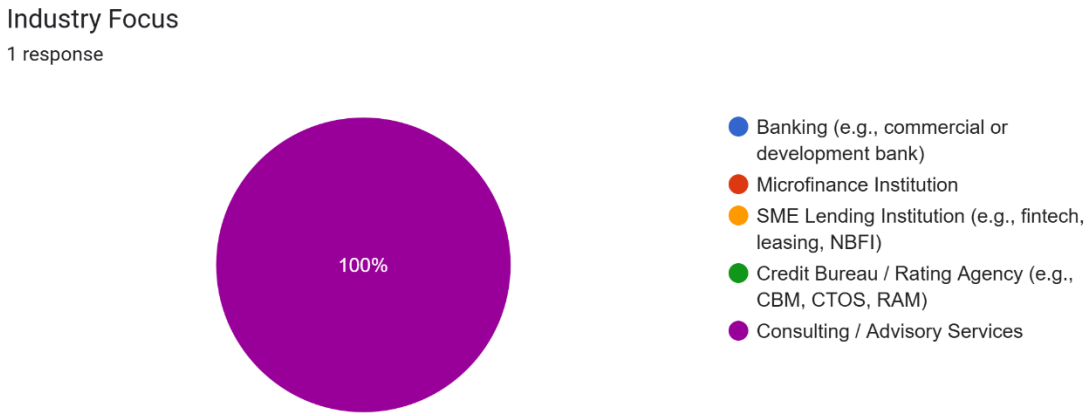


Figure 3.4: Industry Focus of the Expert Respondent

Figure 3.4 highlights that the respondent exclusively operates within Consulting / Advisory Services, as indicated by the full 100% purple segment. Together, both charts provide a clear picture of the respondent's background and support the credibility of the insights gathered through the validation survey.

Indicator Selection Summary

Following the expert profile, the subsequent figures present detailed feedback on indicator usage across five thematic sections, based on the layout of typical credit risk reports. Each chart reflects the expert's selection frequency, providing insight into which indicators are most relied upon during practical MSME credit evaluation.

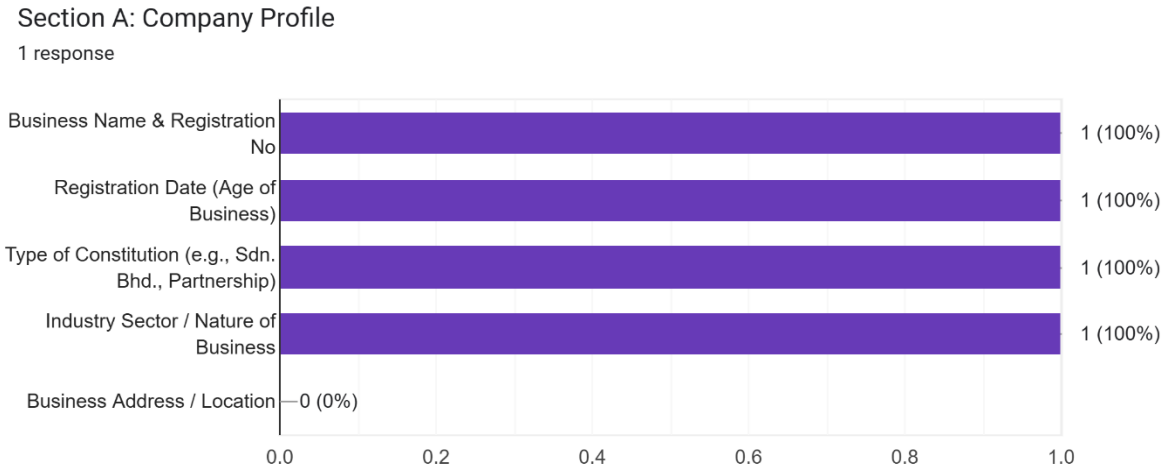


Figure 3.5: Expert Selection of Company Profile Indicators

The first set of indicators, as shown in Figure 3.5, relates to the company's profile. The expert validated four out of five indicators in this category. Specifically, *Business Name & Registration No*, *Registration Date (Age of Business)*, *Type of Constitution*, and *Industry Sector / Nature of Business* were all selected as relevant to creditworthiness evaluation. These details provide foundational context about the business's legal identity, age, and operating environment. In contrast, *Business Address / Location* was not selected, indicating it may carry less weight in decision-making or is considered redundant if industry and registration data are already available. The focus on structural and regulatory details suggests that the expert values legal and sectoral context more than geographic factors during initial assessments.

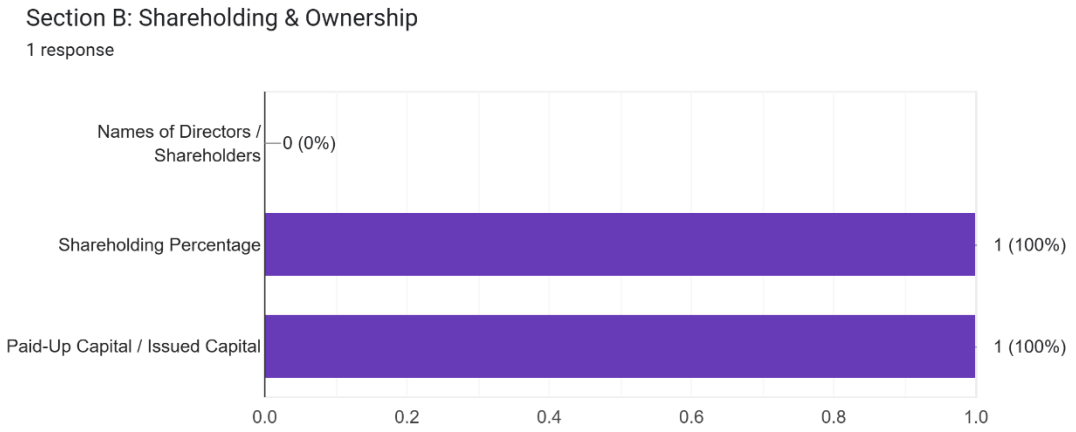


Figure 3.6: Expert Selection of Shareholding and Ownership Indicators

Figure 3.6 presents the expert's selection of indicators related to ownership and company capital. The expert considered *Shareholding Percentage* and *Paid-Up / Issued Capital* as relevant when evaluating MSME creditworthiness. However, they did not select *Names of Directors / Shareholders*. This suggests that the expert places more importance on how ownership is distributed and how much capital has been invested, rather than who the individual owners are. It reflects a practical and objective approach that focuses more on measurable financial data than on personal or reputational factors.

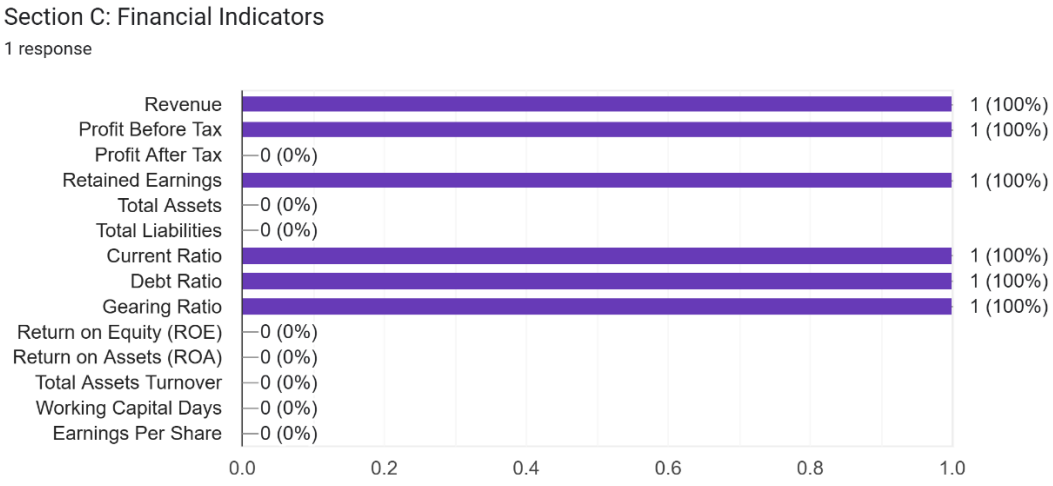


Figure 3.7: Expert Selection of Financial Indicators

In Figure 3.7, the expert selected several core financial indicators: *Revenue*, *Profit Before Tax*, *Retained Earnings*, *Current Ratio*, *Debt Ratio*, and *Gearing Ratio*. These indicators are important because they help assess the company's profitability, financial reserves, liquidity, and level of debt, all of which are critical for understanding whether the business can repay its loans.

The expert did not select other indicators such as *Profit After Tax*, *Total Assets*, *Total Liabilities*, *ROE*, *ROA*, and others. This may indicate that these more complex or derived ratios are either not always available in MSME reports or not as useful in day-to-day evaluation. It also suggests that the expert prioritizes practical, direct measures of performance and risk over more technical or accounting-focused metrics.

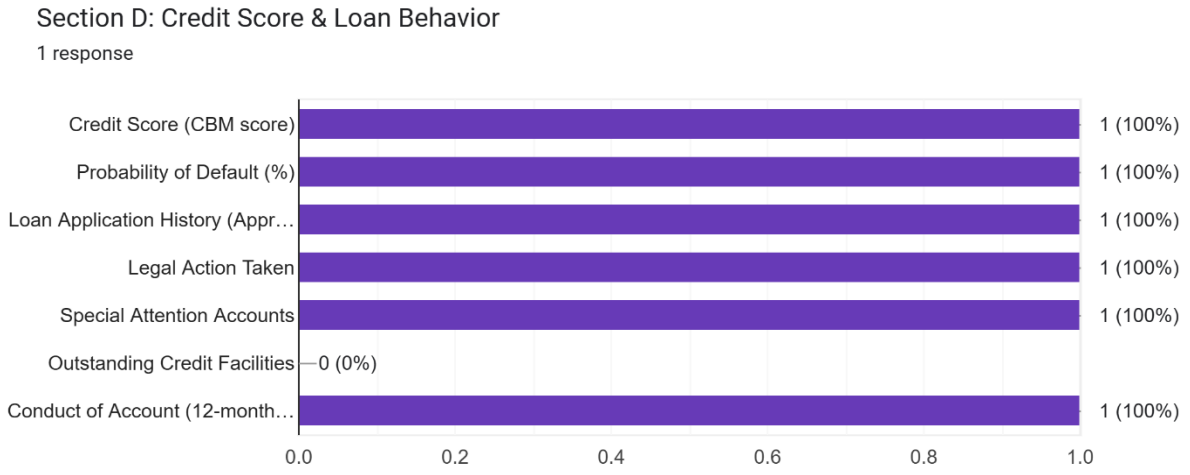


Figure 3.8: Expert Selection of Credit Score and Loan Behaviour Indicators

Figure 3.8 presents a consistent and affirmative validation of almost all credit behaviour indicators. The expert marked *Credit Score (CBM)*, *Probability of Default (%)*, *Loan Application History*, *Legal Action Taken*, *Special Attention Accounts*, and *Conduct of Account (12-month repayment behaviour)* as essential. These indicators directly capture risk exposure, historical conduct, and institutional red flags. Only *Outstanding Credit Facilities* was excluded, which may be due to perceived redundancy if conduct and delinquency patterns already provide sufficient insight into repayment behaviour. This strong emphasis on conduct- and score-based data reinforces the expert's reliance on behavioural indicators in assessing credit risk.

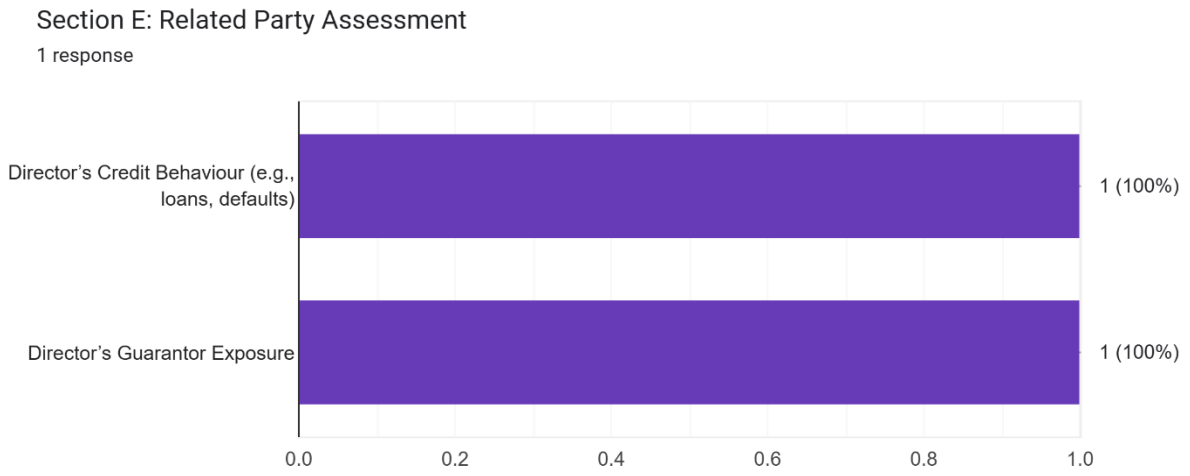


Figure 3.9: Expert Selection of Related Party Assessment Indicators

As seen in Figure 3.9, both *Director's Credit Behaviour* and *Guarantor Exposure* were marked as relevant. This suggests the expert places importance on evaluating not just the business entity, but also the financial conduct and liabilities of individuals closely tied to it. In MSME contexts where business and personal finances often overlap, such checks can provide early warning signals about hidden risks that may not surface through company-level indicators alone.

Additional Suggestions and Comments

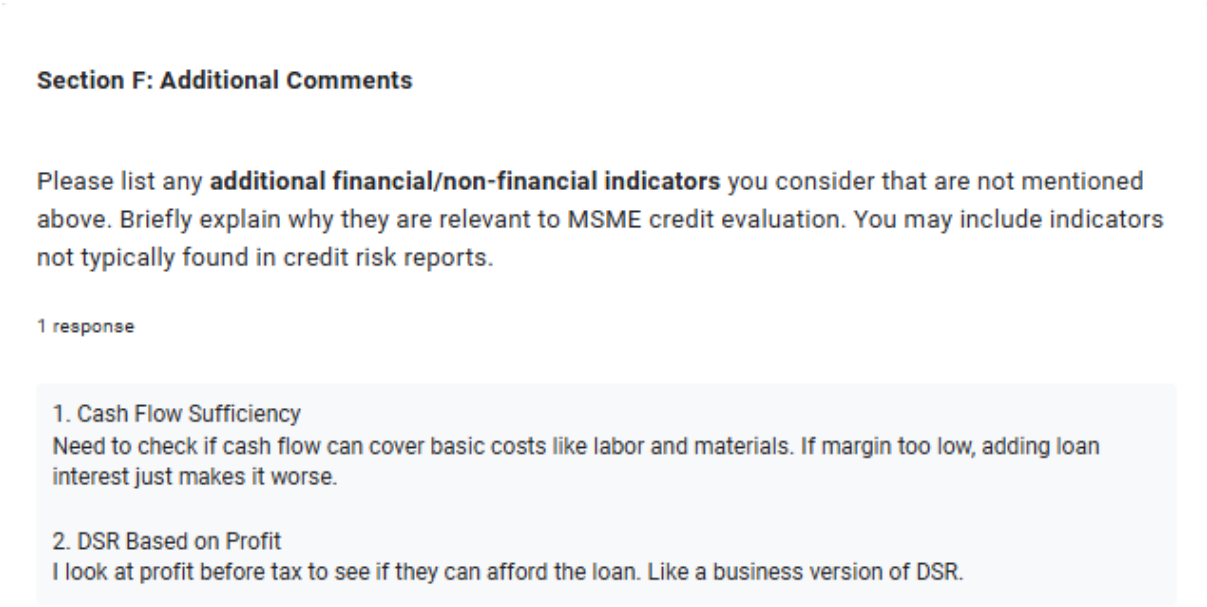


Figure 3.10: Expert Suggestions of Additional Indicators

In the open-ended Section F (Figure 3.10), the expert contributed two additional financial indicators. The first was Cash Flow Sufficiency, with the rationale that cash flow should be able to cover basic operating costs such as labour and materials. The expert noted that in scenarios with low profit margins, additional loan burdens could worsen the business's financial position, thus emphasizing the importance of real liquidity over reported profitability. The second suggestion was DSR Based on Profit, where the expert referred to a business-oriented version of the Debt Service Ratio. Instead of the traditional income-based DSR used in personal finance, the expert looks at Profit Before Tax to determine whether the business can sustain its debt commitments. These suggestions align with practical, real-world evaluations that prioritize affordability and operational resilience.

However, while these indicators are insightful and highly relevant for assessing credit risk, they are not typically included as direct entries in standard credit risk reports. Instead, they must be manually derived using available financial figures, such as profit and expense components, which may require additional computation or contextual interpretation by the credit officer. This highlights a potential gap between what is theoretically useful and what is readily accessible through automated extraction.

Limitations

It is important to note that these findings are exploratory due to the limited number of responses. With only one expert consulted, generalizability is not assumed. Nonetheless, the response provides meaningful practitioner grounding to support and refine the system's feature selection logic. The expert input serves as a valuable intermediary between theory and practice, reinforcing literature-based indicators and ensuring their relevance and feasibility for system implementation.

3.3.3 Feature Selection using AI

To prioritize the most relevant financial indicators for MSME credit evaluation, this study proposes the use of a supervised machine learning technique, Random Forest algorithm, to rank features based on their contribution to credit scoring. The goal is not to develop a deployable predictive model, but rather to identify which indicators are most valuable to extract and present within the rule-based document processing system.

Random Forest is chosen for its robustness with small datasets, tolerance for multicollinearity, and ability to measure feature importance effectively. The algorithm works by constructing an ensemble of decision trees trained on different subsets of the data and averaging their outputs to improve prediction stability. Importantly, it assigns an importance score to each feature based on how much it contributes to reducing prediction error across the ensemble. This ranking helps identify the most influential variables in predicting a target outcome.

In this study, the credit score stated in each credit risk report is used as a proxy for creditworthiness, since explicit approval or rejection decisions are not available. This mirrors common industry practices, where credit scores often serve as a basis for lending decisions.

A dataset of 18 credit risk reports will be used for model training. These reports were selected based on completeness and availability of key information, including credit scores and extractable financial indicators. Two remaining reports are reserved for testing and validation. Prior to modelling, the indicators will be cleaned, standardized, and structured to form three distinct input sets:

- Literature-only feature set: consisting of indicators identified through academic and industry sources,
- Expert-only feature set: derived from the practitioner validation survey,
- Combined feature set: integrating both literature-based and expert-endorsed indicators.

Separating the feature sets into literature-only, expert-only, and combined groups serves an important analytical purpose. It allows for a meaningful comparison between what is recommended in academic research and what is valued in real-world practice. This comparison helps identify indicators that are consistently important across both sources, while also revealing any differences or overlooked insights. By running the model on each set individually, the study can detect whether certain indicators are only supported by theory, only emphasized by practitioners, or validated by both. This process also helps avoid relying too heavily on one perspective and ensures a more balanced, transparent selection.

For each feature set, the Random Forest model will generate a ranked list of indicators according to their importance in predicting credit score. These rankings will inform which features should be prioritized for extraction and display in the system interface. To ensure clarity and relevance, a feature importance threshold will be applied. Since Random Forest generates importance scores ranging from 0 to 1, an initial threshold (e.g., ≥ 0.6) is proposed to filter out weak contributors. This value will be revisited and refined in Capstone Project 2 based on observed score distributions and feature count.

3.3.4 Feature Integration and Filtering

Following the collection of candidate indicators from the literature review and expert validation, and the planned application of Random Forest for feature prioritization, a merging and filtering process will be implemented to finalize the list of extractable features.

Two strategies are considered for integrating feature sets:

- Intersection approach: retains only indicators that appear in both the literature and expert-derived lists. While this ensures strong agreement, it may exclude features that are individually valuable but missed by one source.
- Union approach (preferred): combines all features from both sources and relies on Random Forest to rank them based on empirical importance. This method is more flexible and data-driven, allowing the model to surface high-impact features regardless of source.

The second approach is favoured for this project, as it balances theoretical validity, expert insight, and data-backed relevance. Once merged, the unified feature list will be filtered based on model-generated importance scores, using the pre-set threshold to remove low-ranking indicators. The final output will be a shortlist of N extractable features, which will form the foundation for the rule-based extraction logic implemented in the system.

This strategy combining literature, expert, and AI perspectives ensures that selected indicators are both evidence-based and practically meaningful. The merging and filtering process also ensures that final features are not only important in theory, but also present and extractable in real credit risk report layouts.

3.4 Document Input and Preprocessing

This section outlines the document format used for information extraction and the preprocessing steps applied to prepare the data for structured analysis. The input data consists of MSME credit risk reports issued in PDF format by Credit Bureau Malaysia (CBM), primarily using the standardized MyBizSCoRE layout. Although these reports maintain a relatively consistent structure across borrowers, they still require customized processing logic due to the mix of tables, text blocks, and financial figures. The following subsections detail the format, tools, and extraction strategies employed.

3.4.1 Document Type and Format

The primary document type used in this project consists of MyBizSCoRE credit risk reports provided in PDF format. These reports are semi-structured, containing both structured tables (e.g., financial statements, loan summaries) and free-text sections (e.g., litigation narratives, key influencing factors). The layout is relatively consistent across reports, with key information fields appearing under standardized headers. A total of 20 reports were processed in this study. The document sections and content structure follow the standardized layout introduced in Section 2.3 and a sample of credit risk report is included in Appendix A (Figure A.1, A.2 and A.3).

3.4.2 Justification for Document Focus

While financial institutions may consider a wide range of documents in evaluating MSME creditworthiness, including bank statements, tax records (e.g., SST/GST filings), income statements, and business registration forms, this project deliberately narrows its scope to credit risk reports. The selection is justified by three key factors.

First, **standardization**: Credit risk reports provide a consolidated and structured summary of a borrower's repayment history, outstanding obligations, and historical defaults. Unlike raw financial documents, they present data in a format specifically designed to inform credit decisions.

Second, **automation feasibility**: Credit risk reports are typically semi-structured and machine-readable, which makes them more practical for rule-based or AI-assisted extraction methods. In contrast, documents like scanned tax filings or informal bank statements often vary in format, contain unstructured content, or require extensive manual cleaning, making automated extraction significantly more challenging.

Third, **relevance to digital lending**: Credit risk reports are increasingly adopted by banks and fintech platforms as a core component in MSME credit scoring workflows. Their growing role in automated lending systems enhances both the real-world relevance and future scalability of this project's AI-powered extraction approach.

3.5 Extraction Strategy

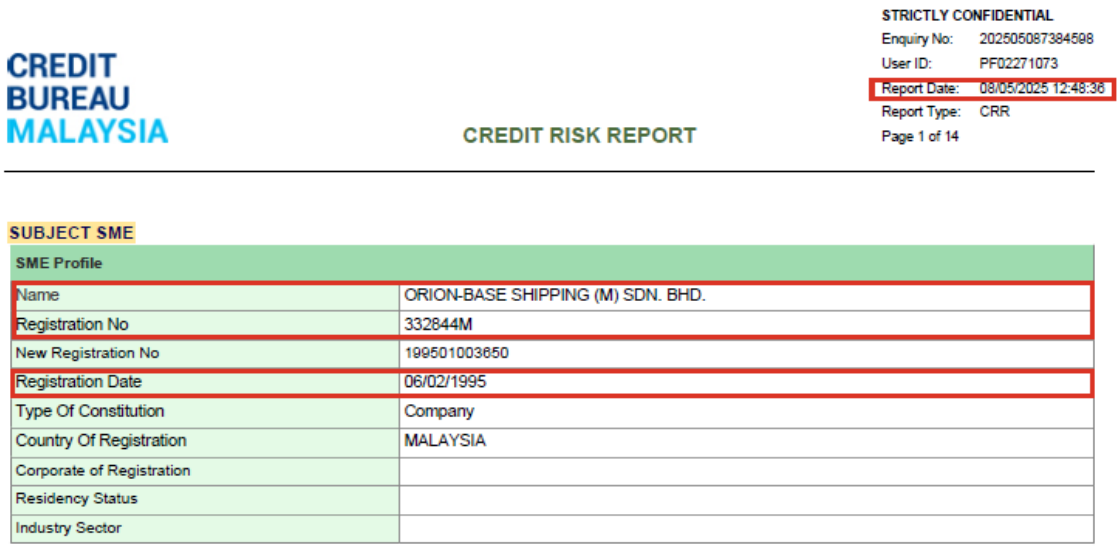
The system uses a modular, rule-based method to extract important data from the credit risk reports. Instead of assuming that each section always appears on the same page, the system looks for specific section titles such as "SUMMARY INFORMATION", "FINANCIAL STATEMENT" or "LOAN INFORMATION", to locate the correct content. This makes the extraction process more reliable when working with reports of varying lengths and layouts.

3.5.1 Company Profile Extraction

At the beginning of each credit risk report, the "SUBJECT SME" section provides basic company metadata. The system extracts only three essential fields from this section:

- Company Name,
- Registration Number
- Registration Date

These fields are selected to support identification, traceability, and derived insights, without overloading the dashboard with low-priority information.



CREDIT BUREAU MALAYSIA

CREDIT RISK REPORT

STRICTLY CONFIDENTIAL
Enquiry No: 202505087384598
User ID: PF02271073
Report Date: 08/05/2025 12:48:36
Report Type: CRR
Page 1 of 14

SUBJECT SME

SME Profile	
Name	ORION-BASE SHIPPING (M) SDN. BHD.
Registration No	332844M
New Registration No	199501003650
Registration Date	06/02/1995
Type Of Constitution	Company
Country Of Registration	MALAYSIA
Corporate of Registration	
Residency Status	
Industry Sector	

Figure 3.11: Business Age is calculated using the registration date and the report issuance date shown in the header.

In particular, the Registration Date is used to compute the company's Business Age, which refers to the number of full years since incorporation. This is a widely used indicator in credit risk analysis, as businesses with longer operational history tend to have more stable financial behaviour, while younger companies are generally prone to volatility or default.

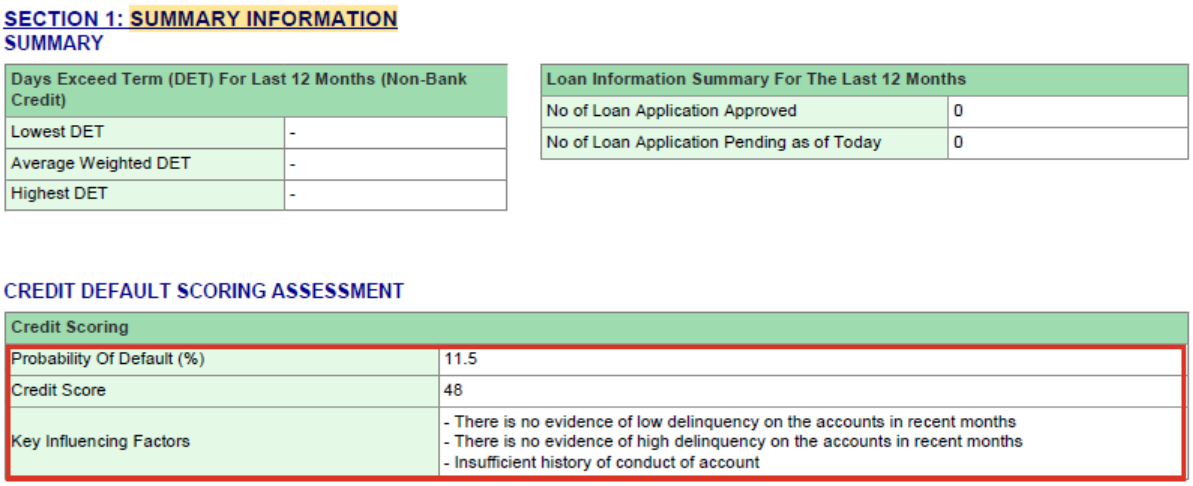
The Business Age is calculated by subtracting the registration date from the credit report's issuance date, which is consistently displayed at the top-right corner of each report page. This ensures that all extracted data points, including conduct history, credit score, and derived metrics like business age, reflect a consistent temporal snapshot. Using the report date instead of the current system date avoids inconsistencies across users and allows reproducible, time-aligned comparisons across multiple reports.

The extracted Company Name, Registration Number, and Business Age are displayed in the main dashboard summary view, allowing credit officers to quickly identify and assess each company alongside financial indicators.

3.5.2 Summary Information

In this section, the key scoring metrics including the values of Probability of Default (%) and the Credit Score are extracted through keyword-based line scanning combined with regular expression matching. This approach allows the system to accurately identify and retrieve

numeric data. In addition, the system captures the list of Key Influencing Factors, which are presented as bullet points, to help users such as credit officers better understand the reasoning behind the assigned credit score.



SECTION 1: SUMMARY INFORMATION

SUMMARY

Days Exceed Term (DET) For Last 12 Months (Non-Bank Credit)	
Lowest DET	-
Average Weighted DET	-
Highest DET	-

Loan Information Summary For The Last 12 Months	
No of Loan Application Approved	0
No of Loan Application Pending as of Today	0

CREDIT DEFAULT SCORING ASSESSMENT

Credit Scoring	
Probability Of Default (%)	11.5
Credit Score	48
Key Influencing Factors	- There is no evidence of low delinquency on the accounts in recent months - There is no evidence of high delinquency on the accounts in recent months - Insufficient history of conduct of account

Figure 3.12: Extracted probability of default, credit score and key influencing factors from the Summary Information section.

3.5.3 Financial Statement

The “FINANCIAL STATEMENT” section of the credit risk report contains key financial data such as balance sheet items, income statement values, and financial ratios. To extract this information, the system uses the pdfplumber.extract_tables() function, which is well-suited for detecting and reading structured tables from machine-readable PDF documents. This allows the system to preserve the original table layout and accurately associate each financial indicator with its corresponding year.

Among the extracted tables, the system identifies the relevant financial ratios table by locating a header row that contains the label “Financial year end.” Once the correct table is found, specific rows of interest, such as Current Ratio (Times) or Profit After Tax Margin (%), are filtered and extracted.

Since not all reports provide data across the same number of years, the system is designed to adapt dynamically. It scans the column headers to detect the financial years available in the report and only extracts values for those years. If fewer years are reported (as opposed to the usual five), the system simply omits the missing ones without error, ensuring consistent processing across both short and multi-year financial tables.

Financial Ratios					
Financial year end	31/12/2022	31/12/2021	31/12/2020	31/12/2019	31/12/2018
Profit Before Tax Margin(%)	0.2	-0.1	-2.8	0.2	-15.5
Profit After Tax Margin(%)	-0.4	-1.1	-0.1	-1.0	-22.3
Current Ratio (Times)	0.4	0.3	0.2	0.2	0.2

Figure 3.13: Extracted financial ratios across multiple years

3.5.4 Loan Information

The "BANKING INFORMATION" section (typically listed as Section 3 in the credit risk report) provides structured data on the company's active credit facilities, including details such as loan type, disbursed amounts, outstanding balances, and repayment conduct over the past 12 months. Although the label "Loan Information" may appear multiple times in the full report, including under related party sections for directors, the system is designed to extract only the loan data belonging to the company itself.

To ensure accuracy, the system first locates the "BANKING INFORMATION" section using stable title-based anchoring rather than section numbers. It then scans only the tables that appear within this section and filters out unrelated content such as director-level banking data found in later sections (e.g., "BUSINESS INTEREST PARTY").

SECTION 3: BANKING INFORMATION

Subject Status	
Warning	-

Figure 3.14: Section 3 header used as an anchor for locating company-level loan data

Once the correct subsection is located, pdfplumber.extract_tables() is used to extract tabular data. The system identifies the loan-level rows by detecting those that begin with numeric loan sequence numbers (e.g., 1, 2, 3) and include associated metadata such as facility type, disbursement date, and conduct history.

For each valid loan, the system extracts the following fields:

- Facility Type (e.g., OTLNFNCE, PCPASCAR)

- 12-Month Conduct of Account
- Legal Status (LGL STS)

Loan Information

No	Date	STS	Capacity	Lender Type	Facility/ App Type	Total Outstanding Balance (RM)	Date Balance Updated	Limit / Instl Amt (RM)	Prin. Repmt. Term	Col Type	Conduct Of Account For Last 12 Months												LGL STS	Date Status Updated
											2025											2024		
Outstanding Credit											M	A	M	F	J	D	N	O	S	A	J	J		
1	07/11/2021		Own	CB				300,000.00																
		O			OTLNFNCE	154,878.00	31/03/2025	6,130.00	MTH	Fin Guarnte			0	0	0	0	0	0	0	0	0			
2	05/07/2023		Own	OWN				10,000.00		Oth Finci Assts														
		O			FNGTRADE	0.00	31/03/2025	0.00	BUL				0	0	0	0	0	0	0	0	0			
3	29/01/2024		Own	OWN				98,100.00		M/Vehs														
		O			PCPASCAR	78,042.00	30/04/2025	1,846.00	MTH			0	0	0	0	0	0	0	0	0	0			
4	30/05/2024		Own	CB				400,000.00		M/Vehs														
		O			PCPASCAR	360,881.00	31/03/2025	7,454.00	MTH				1	1	1	1	1	1	1	0			
5	15/11/2024		Own	CB				10,000.00		Oth Finci Assts														
		O			FNGNTRDE	10,000.00	30/04/2025	0.00	BUL			0	0	0	0	0	0							
		O			OTTRDFAC	0.00	30/04/2025	0.00	BUL			0	0	0										
					Total Outstanding Balance:	603,801.00	Total Limit:	818,100.00																

Figure 3.15: Loan Information section extracted from the company's banking data. Only valid loan rows are processed.

The monthly conduct values are parsed as a list of 12 integers representing repayment behavior, where 0 indicates on-time payment, and higher values (e.g., 3, 5, 15) represent months overdue. Rather than assuming fixed month labels, the system dynamically constructs the correct 12-month window using the report's issuance date, calculated via datetime and dateutil.relativedelta.

To support further risk scoring, two additional conduct-based metrics are computed per loan:

- Max Delinquency: The highest single value in the 12-month conduct history
- Delinquency Count: The number of months with delinquency values ≥ 2

This logic ensures that only the company's banking obligations are extracted, even when personal director-level loan data appears elsewhere in the document.

3.5.5 Special Attention Account and Credit Application

Directly below the loan information table, credit risk reports typically include two additional subsections: Special Attention Account and Credit Application. Although these tables are presented under the broader "Loan Information" heading, they do not contain active or historical loans in the same format as standard facilities, thus they are handled separately here.

The **Special Attention Account** section indicates whether the company has been flagged for accounts requiring extra risk attention (e.g., delinquency, legal action, or judgment). To improve efficiency, the system first checks the “Special Attention Account” field in the **Summary Credit Report**. If the flag is "N" (no account), table-level parsing is skipped. This avoids unnecessary processing of empty structures. However, if the flag is "Y" or if rows are detected in the table below, the system extracts the date, lender, facility type, and asspciated amount. This ensures readiness for future reports where such cases may appear.

Summary Credit Report			
Total No. of Credit Applications			
	No. of Applications	Total Amount (RM)	
A. Approved for past 12 Months	2	770,000.00	
B. Pending	0	0.00	
Summary of Potential & Current Liablities			
	Outstanding (RM) (Exclude FEC)	Total Limt (RM) (Exclude FEC)	FEC Limit (RM)
A. As Borrower	603,801.00	818,100.00	0.00
B. As Guarantor		0.00	0.00
C. Total		818,100.00	0.00
Legal Action Taken	N		
Special Attention Account	N		

Figure 3.16: Special Attention Account status as indicated in the Summary Credit Report section

The Credit Application section lists facilities that have been applied for but not yet accepted or disbursed. While these records do not affect current credit conduct or scoring, they are included to support forward-looking risk analysis, especially when large facility amounts or pending statuses are involved. For each valid row, the system extracts the application date, status (e.g., Pending, Accepted), facility type, and requested loan amount. Property-related fields are excluded due to inconsistent reporting and limited decision relevance.

Credit Application									
1	23/05/2024	T	Own	CB	N			370,000.00	
					PCPASCAR				
	Property Address:								Property Status:
2	30/05/2024	A	Own	CB	N			400,000.00	
					PCPASCAR				
	Property Address:								Property Status:

Figure 3.17: Extracted Credit Application records showing pending or accepted facilities and requested amounts

To maintain a compact dashboard interface, these tables are not shown as part of the main summary. Instead, the extracted records are stored and displayed in expandable sections or tabs when the user clicks to view a specific company profile. This design avoids clutter in the main interface while still offering full transparency when needed.

3.5.7 Excluded Sections

Although certain sample reports include sections such as ‘Trade Credit Reference Information’ and ‘Legal/Winding-Up Litigation Information’, these do not appear in the real credit risk reports used in this study. As such, they are excluded from the extraction pipeline and dashboard design. The system is designed to minimize false assumptions and ensure robustness by emphasizing consistently presented sections.

3.6 Data Cleaning and Post-Processing

Before the data can be used effectively, it is often necessary to be cleaned after the system retrieves the raw values from the credit risk reports. This is because formatting errors or inconsistencies may exist in the extracted data. For example, date values may appear in different formats across reports, and certain numbers may include extra spaces or currency symbols. To address this, a series of cleaning procedures is applied to ensure the data is accurate, consistent, and ready for display or further analysis.

The first step is to eliminate any additional characters that might have been collected during the extraction process, such as leading or trailing spaces, special symbols, or line breaks. The second step converts each value into its appropriate format. A text string such as "RM 1,500.00" is transformed into the numeric value of 1500.00, for instance. In a similar manner, the date "23/01/2021" is converted into a standard format that computers can understand and use.

In some cases, certain indicators may be missing from a report, or the system may not successfully extract them due to variations in layout. When this happens, the missing values are recorded as blank or labeled with a placeholder such as “NaN” (Not a Number) to ensure the data remains well-structured. If the report provides enough related information, the system may attempt to estimate or infer the missing value, such as calculating a total from smaller components.

After cleaning and formatting, the final data is organized into a structured table. In this table, each row represents one credit risk report, and each column corresponds to a specific financial or non-financial indicator, such as registration date, net profit, or loan repayment status. This structured format is stored in a tool called a DataFrame and can also be exported into a CSV file, which is a common spreadsheet format. This makes the cleaned data easy to use for machine learning analysis, expert review, or display on the user interface.

3.7 Dashboard Interface Design

The system includes a proposed interactive dashboard built using Streamlit, an open-source Python framework for building data applications, to promote transparency, usability, and expert validation. The dashboard is designed as the primary interface through which users interact with the system, supporting both document intake and post-extraction review in a unified environment.

The process begins with a document import view, where users can submit credit risk reports in PDF format. Once uploaded, the system automatically processes the file using an predefined extraction pipeline and retrieves key financial and non-financial indicators. Users are then presented with a structured preview of the extracted data, positioned alongside the original document to support verification.

To enhance flexibility, the system offers multiple feature selection modes. Users may choose to extract indicators that are:

- Ranked by AI using a Random Forest model based on their importance in predicting credit score,
- Derived from literature, representing indicators frequently cited in academic or industry research,
- Defined by domain experts based on practical experience,
- Or custom-defined by the user, allowing manual selection or entry (if enabled).

This modular selection system enables users to compare the output generated under different feature sets and supports a semi-automated yet human-controllable workflow for credit risk evaluation.

The extracted results are displayed in a summary dashboard that consolidates key credit indicators across multiple companies. Users may review and edit any of the extracted values directly within the interface to correct potential errors or update information based on expert knowledge. Once validated, the finalized dataset can be exported to CSV or TXT format for integration with modelling tools or downstream systems. This modular and semi-automated design supports a user-guided workflow for credit risk evaluation while remaining adaptable to future extension to broader document types or evaluation criteria.

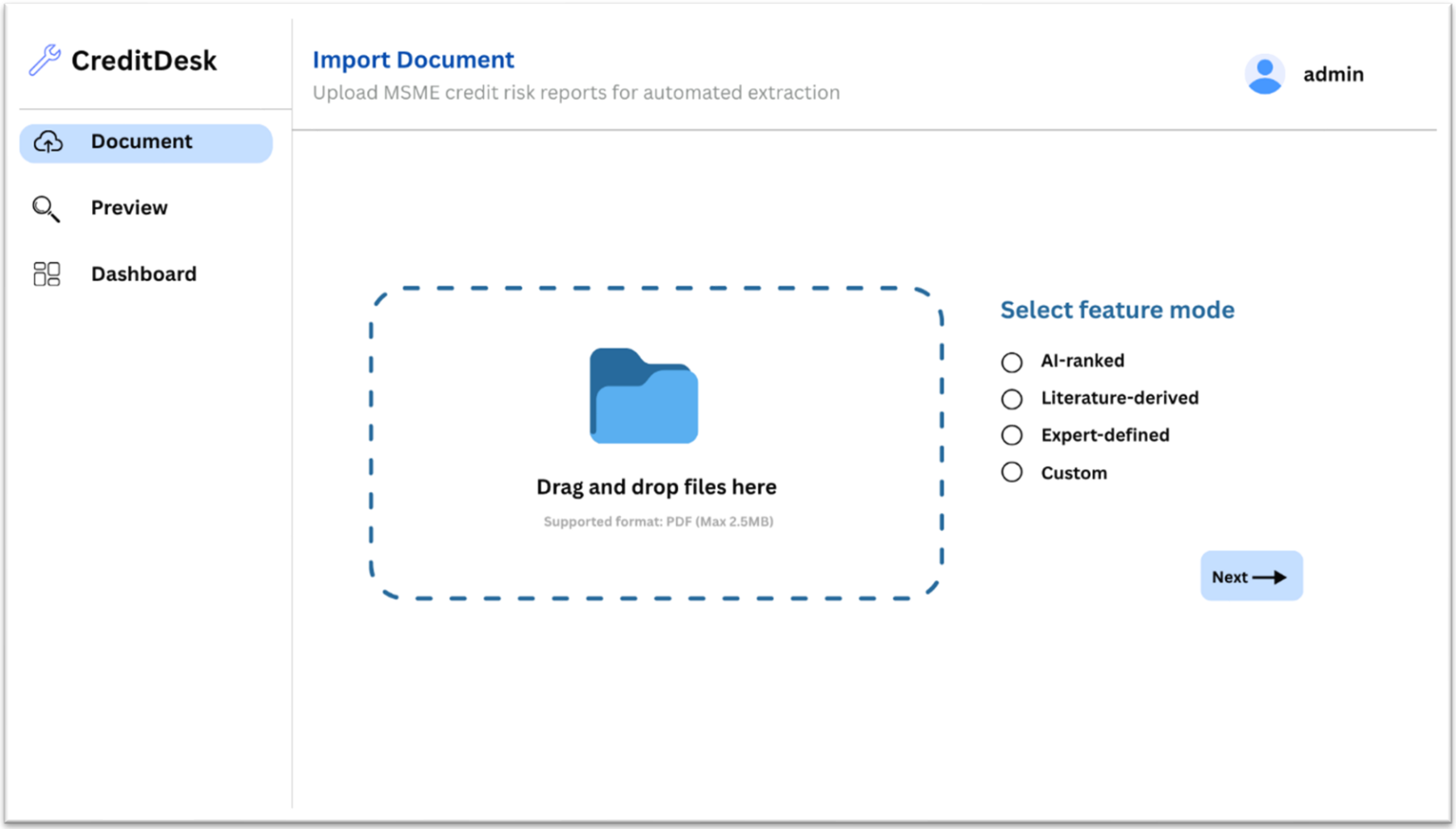


Figure 3.18: Document import and feature selection view

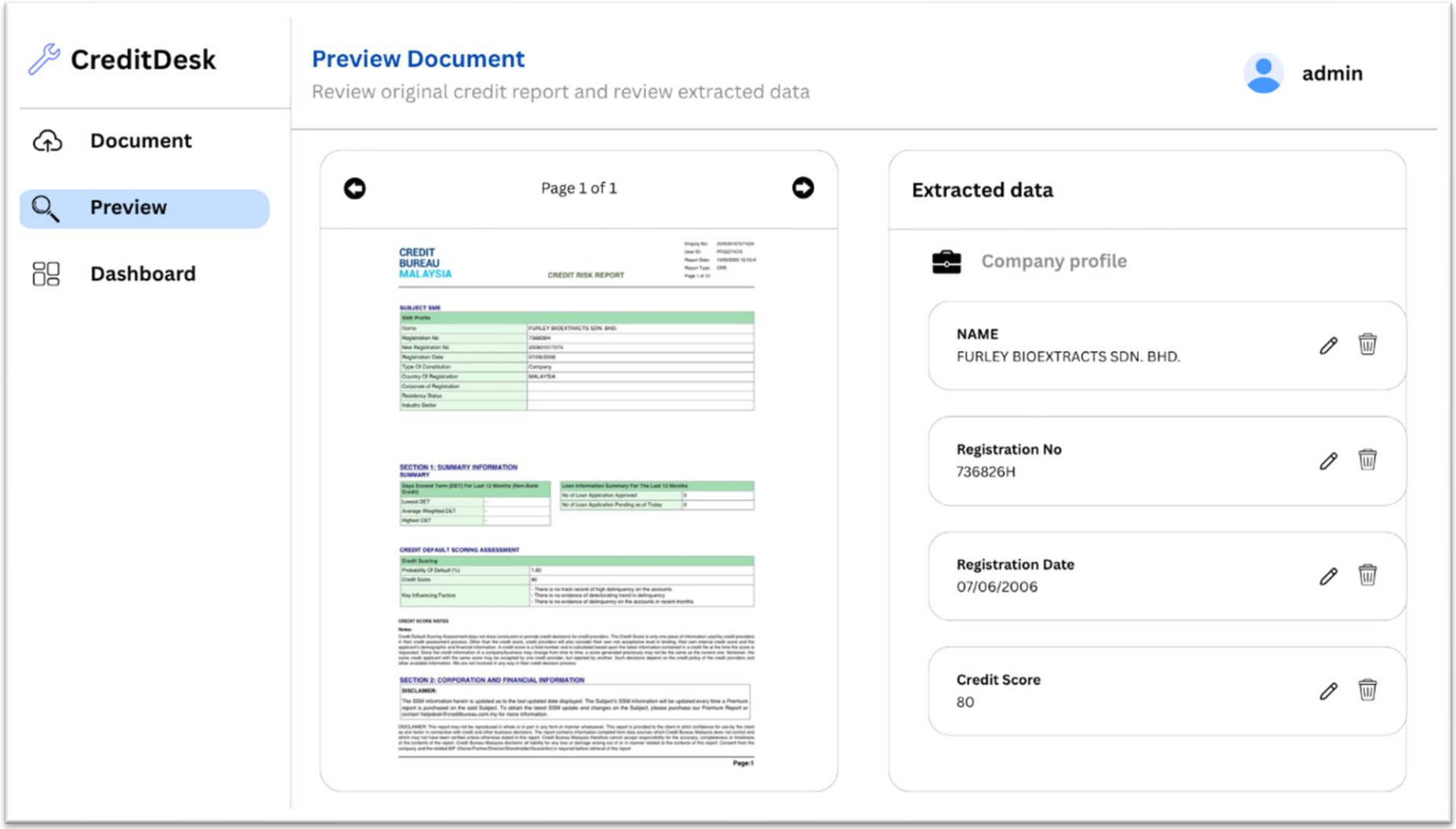


Figure 3.19. Preview interface with extracted data

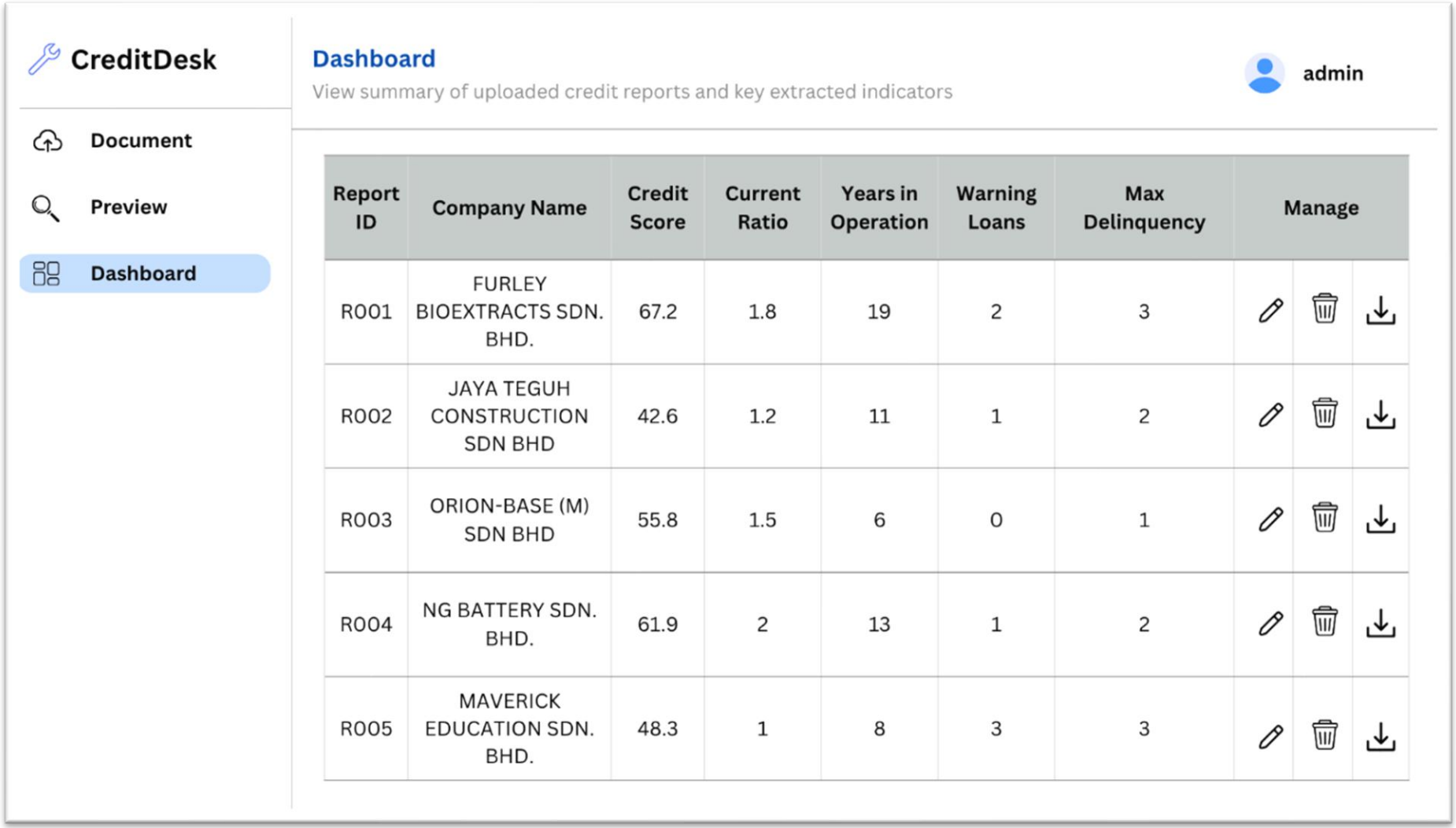


Figure 3.20: Dashboard summary of credit reports and extracted indicator

3.8 System Evaluation

To assess the effectiveness of the proposed AI-powered feature extraction system, a multi-dimensional evaluation approach is applied. This includes benchmarking the system's extraction accuracy, comparing its performance against expert manual effort, and assessing its usability from the perspective of domain users.

3.8.1 Accuracy Benchmarking

The key financial indicators from a representative credit risk report were carefully examined and extracted to create a manually curated ground truth dataset. This dataset serves as the reference baseline for evaluating the correctness of the system's rule-based extraction methods, including techniques such as regular expressions and extract_tables() for structured parsing. After the system automatically processes the same set of reports, its outputs are compared against this ground truth to assess correctness. Every extracted field is assessed as either missing/incorrect, mostly correct, or exact match. The percentage of accurately retrieved fields out of the total predicted is then used to get the overall accuracy score.This quantitative benchmarking provides an objective measure of how reliably the system captures critical information from semi-structured credit documents.

Table 3.2: Extraction Accuracy Comparison Between Ground Truth and System

Report ID	Indicator	Ground Truth Value	System Extracted Value	Accuracy Status
R01	Current Ratio (2022)	1.25	1.25	✓
R01	Current Ratio (2021)	0.85	0.58	×
R01	Profit After Tax Margin (%)	-0.4	0.4	!
R01	Credit Score	48	48	✓
R01	Total Credit Limit	RM 800,000	—	×

Note: Table 3.2 presents selected indicators for illustrative purposes only.

Legend for Accuracy Status:

- Exact Match (✓): System output matches ground truth exactly (format + value).
- Partial Match (!): Slight variation in format or precision, but semantically correct.
- Incorrect/Missing (✕): System missed or extracted the wrong value.

Table 3.3: Summary Table

Total Fields Evaluated	Exact Matches	Partial Matches	Incorrect/Missing	Overall Accuracy (%)
20	17	2	1	85%

Note: The full evaluation of 20 extracted fields from a single representative credit risk report is reflected in the summary statistics in Table 3.3.

3.8.2 Performance Comparison with Experts

To further validate the system's practical value, a task-based performance comparison was carried out between the AI-assisted system and human experts. Financial indicators were manually extracted from the same credit risk report used for the system evaluation by a chosen group of domain experts.

For each expert session, two key metrics are used:

(i) the time taken to complete the extraction task (reflecting efficiency)

(ii) the accuracy of their results, benchmarked against the manually curated ground truth (as described in Section 3.8.1).

A structured results table was then used to compare the system's performance against the human baseline. This enabled an assessment of whether the AI-assisted system could deliver meaningful time savings without compromising accuracy, thereby supporting its integration into professional credit review processes.

A comparative table will be used to summarize performance outcomes, highlighting differences in time efficiency and accuracy between the system and expert users.

Table 3.4: Performance Comparison Between System and Human Experts

Participant	Extraction Mode	Total Time Taken (min)	Accuracy (%)	Notes
Expert 1	Manual	14.2	90%	Missed 2 values
Expert 2	Manual	17.5	85%	Slower due to unfamiliar layout
Expert 3	Manual	12.8	90%	Few format inconsistencies
System	Automated	**2.1**	**95%**	One partial match

3.8.3 Usability and Feature Relevance Assessment

The system's usability was assessed, alongside its technical performance, to ensure it was appropriate for practical use in institutional environments. Following their interaction with the system, a post-task survey was administered to expert users to capture their perceptions of its ease of use and overall acceptance.

The Technology Acceptance Model (TAM), a well-known paradigm for determining users' propensity to accept new technologies, served as the foundation for the survey's design. It emphasizes two core ideas: Perceived Usefulness (PU), the extent to which a system improves task performance, and Perceived Ease of Use (PEOU), the degree to which a system is simple and intuitive to operate. The TAM literature provides strong evidence for these constructs and their impact on user intention (Ozag & Duguma, 2004; Immanuel & Thoyyibah, 2025).

In accordance with TAM2, the questionnaire has been expanded to offer a more thorough understanding of user approval (Venkatesh & Davis, 2000). Two other dimensions were included: Attitude Toward Use (ATU) and Behavioural Intention to Use (BI). These additions are further supported by Ozag and Duguma (2004), who explain that users' beliefs about a system's usefulness (PU) and their intention to use it (BI) are influenced by more than just system design. They are also shaped by the way users perceive the system's functionality, how autonomous they feel when using it, and the rationale behind their decision to adopt it.

The final instrument consisted of 13 close-ended items, grouped into four sections:

- Section A: Perceived Usefulness, which focuses on aspects such as task efficiency and the relevance of extracted indicators.
- Section B: Perceived Ease of Use, which covers factors like ease of learning, interface intuitiveness, simplicity of interaction.
- Section C: Attitude Toward Use, which assesses user satisfaction and preference.
- Section D: Behavioural Intention, which evaluates the likelihood of reuse or recommendation to others.

A 5-point Likert scale was used to rate each item, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). An open-ended section was further adds on to gather qualitative feedback and suggestions for improvement.

Notably, one of the items under the *Perceived Usefulness* section (Item 3) asked participants to assess whether the extracted indicators were relevant and useful for MSME credit assessment. Given that Random Forest analysis was used to prioritize the selected indicators, this serves as an indirect validation of the system's feature selection methodology. Even when the underlying model is not explicitly referenced, positive responses to this item indicate expert alignment with the system's chosen features.

A purposive sample of domain specialists, such as advisors, financial analysts, or credit officers, received the survey. Participants were instructed to simulate real-world system usage before responding to the survey, matching with Immanuel and Thoyyibah (2025) task-based evaluation design. As a result, their responses reflected authentic interaction experiences and provided actionable insights for future system refinements and deployment.

Since the assessment needed feedback from people with relevant domain expertise and experience with credit assessment workflows, the survey was distributed directly to chosen participants via email and professional messaging platforms using a protected Google Form.Participants can access the form using the following link: https://docs.google.com/forms/d/e/1FAIpQLSdC5enlGowUjw-BcRE5r_9pVzDkN-TLrRER99ooB0dLjey02w/viewform.

3.9 Tools and Environment

A collection of tools and technologies, selected for their accessibility, adaptability, and alignment with the project's goals, was used in the development and deployment of the proposed system. The tools fall into two main categories: programming language and supporting libraries.

Programming Language:

Python was chosen as the primary programming language due to its simplicity, robust ecosystem, and extensive support for data processing and machine learning. Its widespread use in both academic and industrial settings ensures compatibility with various open-source tools and community-driven development.

Libraries and Packages:

Several Python libraries were utilized to support different components of the system:

- PDFPlumber: This library enables precise extraction of text, tables, and metadata from scanned or digitally generated PDF credit risk reports. It was used to implement the rule-based extraction logic across various sections of the documents.
- Pandas: Used for efficient data manipulation, cleaning, and transformation. Pandas DataFrames formed the core data structure for storing and processing extracted financial indicators.
- Scikit-learn: This package provided the implementation of the Random Forest model used for feature importance ranking. It also supported essential machine learning functions such as train-test splitting, model evaluation, and result interpretation.
- Streamlit: A lightweight open-source framework used to build the interactive dashboard for file uploads, data visualization, manual review, and CSV export. Streamlit enabled the creation of a user-friendly interface suitable for non-technical stakeholders such as credit officers.

These tools were chosen for their ease of integration and ability to support rapid development of functional prototypes in academic projects.

4.0 Work Plan and Timeline

This chapter outlines the comprehensive work plan, project timeline, and risk mitigation for the Final Year Project, which is organized into two distinct phases: Capstone Project 1 (CP1) and Capstone Project 2 (CP2). Phase spans 14 weeks, while phase 2 requires 12 weeks, amounting to a total project duration of 26 weeks. The planning and research components of the project, such as problem understanding, literature review, feature identification, methodology development, and work schedule preparation, are the primary focus of CP1. The system implementation and evaluation activities in CP2 will be based on the outcomes of this phase.

The second phase, Capstone Project 2, encompasses the development, testing, and refinement of the proposed system. This includes building the rule-based extraction module, developing the Streamlit-based dashboard interface, and applying the Random Forest model to rank financial indicators based on importance. The final results are then validated through expert feedback and usability testing.

Across both phases, the project will produce the following key deliverables:

1. Planning documents
2. Activity log
3. Annotated bibliography
4. Validated list of financial indicators for MSME credit evaluation
5. Feature ranking output produced by Random Forest model
6. Python script for rule-based extraction of structured financial data from credit risk reports
7. Interactive dashboard for report upload, data review, editing, and CSV export
8. TAM-based usability evaluation results

A detailed Gantt chart is also included to visualize the progression and timeline of each task. This chapter concludes with a risk assessment table that outlines potential project challenges and proposed mitigation strategies.

4.1. Capstone Project Work Breakdown Plan

The task breakdown and timeline outlined in this section provide a structured view of all activities planned for the capstone project, from initiation to final submission. It details each task's duration, description, expected outputs, potential risks, dependencies, and corresponding mitigation strategies, ensuring that the project is executed in a controlled and well-monitored manner. This structured approach not only clarifies the sequence of work but also supports effective tracking of progress and timely identification of potential issues. By following this plan, both phases of the capstone project, Capstone Project 1 (planning, research, and preliminary development) and Capstone Project 2 (execution, evaluation, and closure), are aligned towards achieving the defined objectives within the allocated timeframe.

Table 4.1 Task Breakdown and Timeline for Capstone Project

No	Task	Duration	Description	Resultant Work	Risk Factors	Dependencies	Mitigation Strategy
1.0	**Initiation Phase**						
1.1	Join Capstone Briefing Session	1 day	Attend the scheduled CP1 briefing session delivered by Capstone Coordinator, Ms. Lim Woan Ning.	Notes capturing CP1 structure, timeline, and supervisor expectations	Overlooking important briefing points.	-	Review slides or recording of the briefing.
1.2	Explore Feasible Research Ideas	10 days	Conduct early research on topic of interest.	List of viable project topics	Difficulty identifying suitable topic.	1.1	Seek suggestions from peers and supervisor.
1.3	Conduct literature review on the interested topic	7 days	Review and analyse existing literature relevant to the chosen research topic.	Summary of key findings and research gaps.	Selecting irrelevant sources.	1.2	Use reputable sources and record them in an Excel-based annotated bibliography.

1.4	Preliminary Research Discussion with Supervisor	3 days	Meet with supervisor to discuss project's research area, scope and feasibility.	Academic feedback gained to refine project direction and topic	Misalignment with supervisor's expectations causing rework	1.3	Record all feedback to ensure accurate project refinement
1.5	Confirm Project Title with Supervisor	4 days	Discuss and confirm final topic based on preliminary findings.	Approval for final year project topic	Topic not accepted by supervisor.	1.4	Prepare multiple backup topic ideas.
1.6	Submit SAF form	2 days	Submit completed SAF with required details and signatures.	Signed SAF confirming project roles and responsibilities.	Missing project details leading to form rejection	1.3, 1.4	Review form with supervisor ahead of the deadline to ensure accuracy
MILESTONE: FINALIZED PROJECT TITLE							-
1.7	Prepare Initial Project Plan	5 days	Draft the initial project plan including objectives, scope, activities, milestones, deliverables and timeframe based on the approved project topic.	Completed Initial Project Plan, Gantt Chart and draft and introduction.	Missing key tasks or inaccurate time estimates.	1.5, 1.6	Consult the supervisor to evaluate the project timeline and feasibility with the given timeframe.
2.0	**Planning Phase**						

2.1	Gather Relevant Literature	6 days	Collect academic and industry papers on proposed topic	Annotated literature review compiled in Excel format	Incomplete or unfocused reading.	1.7	Use Mendeley to organize papers.
2.2	Draft Structure for Chapter 2	2 days	Determine structure of Chapter 2 based on themes, timeline, or methods	Structured outline of Chapter 2	Limited depth into relevant themes and research areas	2.1	Review past papers and academic examples to guide literature structure
2.3	Create Literature Table	4 days	Compile key sources with titles, area addressed, methods and findings.	Completed literature table.	Missing critical studies.	2.1	Continuously update table as new relevant sources are identified.
2.4	Build Literature Comparison Table	4 days	Identify innovations, recurring patters and gaps between studies.	Comparative summary table of academic sources	Insufficient comparison leading to weak justification of project novelty	2.2, 2.3	Perform second-level review of key sources
2.5	Enhance Structure of Chapter 2	2 days	Sequence subtopics and plan transitions.	Logical flow of Chapter 2	Unclear transitions.	2.2, 2.4	Use mind map or thematic grouping.
2.6	Draft Chapter 2	14 days	Write full literature review.	Completed Chapter 2 draft	Lack of depth or cohesion.	2.5	Seek iterative feedback from supervisor.
2.7	Review previous studies methodology	7 days	Analyse research methodologies applied in previous related studies to identify best practices and common approaches.	Summary of methodologies with strengths and limitations.	Misinterpreting methods or overlooking relevant studies.	2.4, 2.6	Cross-check methodology details with multiple sources and verify through

							supervisor feedback.
2.8	Plan Overall Methodology	3 days	Create a detailed plan explaining both the techniques and workflow used in project	Overview draft for Chapter 3.	Ambiguity in chosen approach.	2.7	Reference methodology chapters in related works.
2.9	Define Feature Selection Strategy	3 days	Design the merging, validation, and Random Forest ranking approach.	Feature selection method section.	Incomplete or biased selection logic.	2.7, 2.8	Validate approach with supervisor and expert feedback.
2.10	Define Extraction Logic	3 days	Explain rule-based extraction such as PDF anchors and table parsing.	Extraction method section.	Ambiguity in rule logic.	2.7, 2.8	Test logic on sample PDFs.
2.11	Outline Evaluation Method	3 days	Plan usability testing and performance metrics.	Evaluation section draft	Evaluation plans unclear or infeasible.	2.8, 2.9, 2.10	Review TAM examples or testing methods.
2.12	Prepare completed Chapter 1 introduction	5 days	Write full version based on refined structure.	Finalized Chapter 1	Disjointed or unclear transitions between subsections.	1.7, 2.6, 2.8	Follow recommended format and get feedback.
	Final Documentation						
2.13	Complete Citation and Reference List	2 days	Ensure citations and references follow proper format in APA7	Final reference list	Citation formatting issues.	2.6, 2.12	Use referencing tool and double check style.
2.14	Finalize Planning Document	4 days	Review and polish entire CP1 document.	Finalized planning document	Formatting or missing content.	2.12	Proofread in multiple rounds.

2.15	Compile Supervision Meeting Records	2 days	Complete all supervision meeting logs and send to supervisor for signature.	Completed meeting record.	Missing details from meetings.	-	Refer to handwritten/typed notes from meetings.
2.16	Finalize Activity Log	2 days	Compile all the necessary documents to be included in the Activity Log,	Updated Activity Log including Supervision Meeting Records and Annotated Bibliography.	Inconsistent documentation.	-	Cross-check with original timeline and edits.
2.17	Submit CP1 Document	1 day	Submit Planning Document and Activity Log through official channel.	Proof of submission.	Technical submission failure.	2.13, 2.14, 2.15, 2.16	Submit early and verify receipt.
MILESTONE: SUBMISSION OF CAPSTONE PROJECT 1							-
3.0	**Execution Phase**						
Step 1: Data Preparation and Feature Selection							
3.1	Collect finalised sample credit risk reports	3 days	Gather complete set of reports for system testing	Final report dataset	Incomplete or inconsistent reports	-	Verify completeness and format before processing
3.2	Prepare dataset of extracted indicators for model input	5 days	Structure and format indicators for ML use	Clean, formatted dataset	Incorrect data mapping	3.1	Cross-check mapping with source reports

3.3	Train Random Forest model for feature importance ranking	6 days	Run model to rank financial indicators	Feature importance list	Overfitting or biased ranking	3.2	Use cross-validation and balanced datasets
3.4	Evaluate ranking results against expert feedback and literature	5 days	Compare rankings with validated sources	Validated feature ranking	Conflicting results with experts	3.3	Reassess parameters and consult experts
3.5	Finalise prioritised feature set	2 days	Confirm final list of indicators for system	Approved feature list	Inclusion of low-relevance features	3.4	Filter using dual validation
Step 2: Extraction Logic & Data Handling							
3.6	Implement section-specific extraction logic	8 days	Code rules for extracting indicators by report section	Functional extraction Python-based scripts	Misaligned text anchors or parsing errors	3.5	Test with varied report formats
3.7	Validate extraction results against ground truth dataset	5 days	Compare outputs to manually verified data	Accuracy report	Low extraction accuracy	3.6	Adjust parsing rules
3.8	Data cleaning and post-processing	4 days	Format, normalise, and remove data inconsistencies	Clean, analysis-ready dataset	Loss of important data during cleaning	3.7	Keep backup of raw extraction
Step 3: System Development							
3.9	Implement Streamlit-based dashboard	7 days	Build interactive UI for data display and editing	Functional dashboard interface	Poor usability	3.8	Gather user feedback during development
3.10	Integrate CSV export functionality	3 days	Enable export of processed data to CSV	Working export feature	Incorrect data format on export	3.9	Validate export output
3.11	Implement data visualisations	5 days	Add summary tables, delinquency highlights, and charts	Visual analytics module	Misleading visuals	3.9	Match visuals to verified data

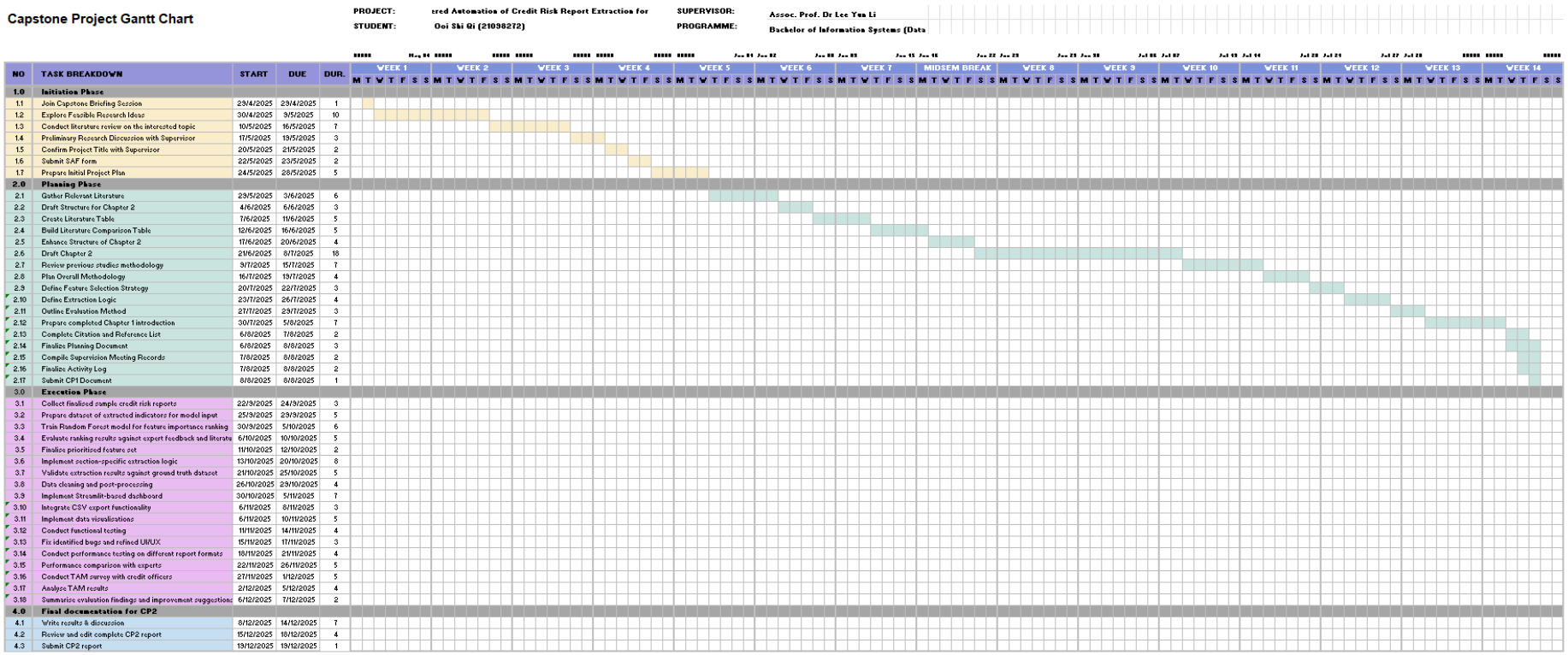
Step 4: Testing and Optimisation							
3.12	Conduct functional testing	4 days	Test all features for intended behaviour	Functional test report	Missed test cases	3.10, 3.11	Use a detailed test checklist
3.13	Fix identified bugs and refined UI/UX	3 days	Resolve issues found in testing	Improved, stable system	New bugs after fixes	3.12	Regression testing
3.14	Conduct performance testing on different report formats	4 days	Assess speed and accuracy across file types	Performance evaluation report	Poor performance on certain formats	3.13	Optimise preprocessing for problematic formats
Step 5: Evaluation							
3.15	Performance comparison with experts	5 days	Performance comparison with experts	Comparative performance report	Significant gaps between results	3.14	Analyse gaps and adjust logic
3.16	Conduct TAM survey with credit officers	5 days	Gather feedback on system usability and usefulness	Completed TAM survey	Low response rate	3.15	Provide clear instructions and follow-ups
3.17	Analyse TAM results	4 days	Process survey responses into insights	TAM analysis report	Misinterpretation of data	3.16	Use statistical validation
3.18	Summarise evaluation findings and improvement suggestions	2 days	Compile results into actionable recommendations	Evaluation summary document	Missing critical improvement points	3.17	Review with experts
4.0	**Final documentation for CP2**						
4.1	Write results & discussion	7 days	Summarise and analyse system performance, feature ranking, and TAM findings	Results & discussion chapter	Weak linkage to objectives	3.18	Align discussion with research goals

4.2	Review and edit complete CP2 report	4 days	Proofread and refine full report content	Finalised report draft	Lingering grammatical or formatting issues	4.1	Multiple review passes and peer review
4.3	Submit CP2 report	1 day	Deliver report through official channel	Proof of submission	Late or failed submission	4.2	Submit early and confirm receipt
MILESTONE: SUBMISSION OF CAPSTONE PROJECT 2							

4.2 Gantt Chart for Capstone Project

The Gantt chart presented in this section offers an at-a-glance view of the planned timeline, showing how the project will progress from start to finish across both phases of the capstone. It maps out the sequence, duration, and overlap of major activities, helping to visualise how each stage connects to the next over the full 26-week period. The first phase, Capstone Project 1, focuses on planning, research, and preliminary development, as illustrated in Figure 4.1 below. The second phase (figure 4.2) will then move into the execution, monitoring, and closure stages, bringing the proposed project to completion. The complete Gantt chart can be accessed through the following link:

https://docs.google.com/spreadsheets/d/1jQB0Q6QhK6tg_1mSOY0FHNVno-JwcYdN/edit?usp=sharing&ouid=114346025408469753813&rtpof=true&sd=true



NO	TASK BREAKDOWN	START	DUE	DUR.
1.0	**Initiation Phase**			
1.1	Join Capstone Briefing Session	29/4/2025	29/4/2025	1
1.2	Explore Feasible Research Ideas	30/4/2025	9/5/2025	10
1.3	Conduct literature review on the interested topic	10/5/2025	16/5/2025	7
1.4	Preliminary Research Discussion with Supervisor	17/5/2025	19/5/2025	3
1.5	Confirm Project Title with Supervisor	20/5/2025	21/5/2025	2
1.6	Submit SAF form	22/5/2025	23/5/2025	2
1.7	Prepare Initial Project Plan	24/5/2025	28/5/2025	5
2.0	**Planning Phase**			
2.1	Gather Relevant Literature	29/5/2025	3/6/2025	6
2.2	Draft Structure for Chapter 2	4/6/2025	6/6/2025	3
2.3	Create Literature Table	7/6/2025	11/6/2025	5
2.4	Build Literature Comparison Table	12/6/2025	16/6/2025	5
2.5	Enhance Structure of Chapter 2	17/6/2025	20/6/2025	4
2.6	Draft Chapter 2	21/6/2025	8/7/2025	18
2.7	Review previous studies methodology	9/7/2025	15/7/2025	7
2.8	Plan Overall Methodology	16/7/2025	19/7/2025	4
2.9	Define Feature Selection Strategy	20/7/2025	22/7/2025	3
2.10	Define Extraction Logic	23/7/2025	26/7/2025	4
2.11	Outline Evaluation Method	27/7/2025	29/7/2025	3
2.12	Prepare completed Chapter 1 introduction	30/7/2025	5/8/2025	7
2.13	Complete Citation and Reference List	6/8/2025	7/8/2025	2
2.14	Finalize Planning Document	6/8/2025	8/8/2025	3
2.15	Compile Supervision Meeting Records	7/8/2025	8/8/2025	2
2.16	Finalize Activity Log	7/8/2025	8/8/2025	2
2.17	Submit CP1 Document	8/8/2025	8/8/2025	1
3.0	**Execution Phase**			
3.1	Collect finalised sample credit risk reports	22/9/2025	24/9/2025	3
3.2	Prepare dataset of extracted indicators for model input	25/9/2025	29/9/2025	5
3.3	Train Random Forest model for feature importance ranking	30/9/2025	5/10/2025	6
3.4	Evaluate ranking results against expert feedback and literatu	6/10/2025	10/10/2025	5
3.5	Finalize prioritized feature set	11/10/2025	12/10/2025	2
3.6	Implement section-specific extraction logic	13/10/2025	20/10/2025	8
3.7	Validate extraction results against ground truth dataset	21/10/2025	25/10/2025	5
3.8	Data cleaning and post-processing	26/10/2025	29/10/2025	4
3.9	Implement Streamlit-based dashboard	30/10/2025	5/11/2025	7
3.10	Integrate CSV export functionality	6/11/2025	8/11/2025	3
3.11	Implement data visualisations	6/11/2025	10/11/2025	5
3.12	Conduct functional testing	11/11/2025	14/11/2025	4
3.13	Fix identified bugs and refined UI/UX	15/11/2025	17/11/2025	3
3.14	Conduct performance testing on different report formats	18/11/2025	21/11/2025	4
3.15	Performance comparison with experts	22/11/2025	26/11/2025	5
3.16	Conduct TAM survey with credit officers	27/11/2025	1/12/2025	5
3.17	Analyse TAM results	2/12/2025	5/12/2025	4
3.18	Summarise evaluation findings and improvement suggestions	6/12/2025	7/12/2025	2
4.0	**Final documentation for CP2**			
4.1	Write results & discussion	8/12/2025	14/12/2025	7
4.2	Review and edit complete CP2 report	15/12/2025	18/12/2025	4
4.3	Submit CP2 report	19/12/2025	19/12/2025	1

Figure 4.1: Gantt Chart for Capstone Project (Phase 1)

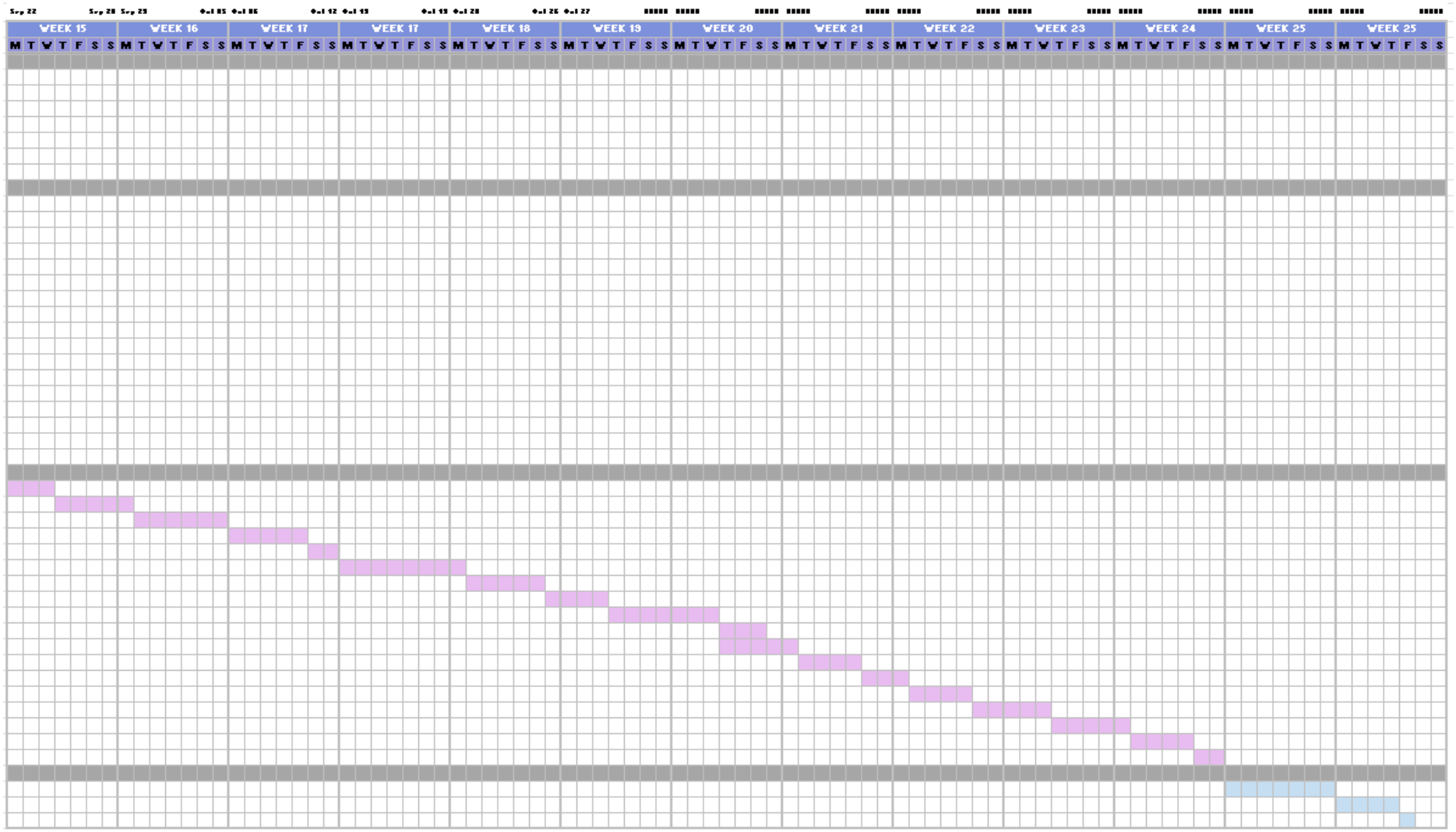


Figure 4.2: Gantt Chart for Capstone Project (Phase 2)

4.3 Risk Assessment

This section analyses and assesses potential risks that may affect the successful completion of both Capstone Project 1 and Capstone Project 2. Table 4.3 lists these risks, which are evaluated based on their likelihood and potential impact, with corresponding mitigation strategies proposed to reduce their effects. The risks cover technological, procedural, and user-related issues that may arise throughout the project's lifecycle. By anticipating these challenges and implementing proactive measures, the project can maintain steady progress, ensure quality, and achieve the specified objectives within the planned timeframe.

Table 4.1: Risk Assessment Table

Risk	Risk Level	Impact of Risk	Mitigation Strategy
Extraction Logic Errors	High	Incorrect or incomplete extraction of financial indicators may reduce system accuracy and undermine result reliability.	Conduct multiple testing rounds on diverse report samples; manually verify extracted values; refine and adjust rule-based logic where errors occur.
Limited Expert Availability	Medium	Delays in obtaining expert feedback for indicator validation may slow finalising the feature set and decision-making.	Schedule consultations well in advance; prepare concise, structured questions; identify backup domain experts.
Schedule Overruns in CP2	Medium	Extended testing or evaluation phases may reduce available time for documentation, impacting report quality.	Monitor progress closely; set internal deadlines ahead of official ones; reallocate resources to critical tasks.
User Acceptance Issues	Medium	Credit officers may be reluctant to adopt the system if perceived as complex or unreliable, affecting real-world applicability.	Conduct early usability testing; incorporate user feedback; design an intuitive and user-friendly interface.
Unforeseen Technical Errors in Deployment	Medium	Bugs arising during live testing could delay evaluation and final submission.	Implement staged deployment; maintain a detailed bug log; allocate buffer time for troubleshooting and fixes.

References

Abu Hassan, A. (2020). *Small and medium enterprise (SME) financing: Credit evaluation method*. In *FBM Insights* (Vol. 2, pp. 1–3). Universiti Teknologi MARA (UiTM) Cawangan Kedah. https://ir.uitm.edu.my/id/eprint/49582/

Al-Slehat, Z. A. F., Almanaseer, S. R., Al Sharif, B. M. M., Al-Haraisa, Y. E., Aloshaibat, S. D., & Almahasneh, M. A. (2024). Creditworthiness Criteria According to the 5Cs Model and Credit Decision: The Moderating Role of Intellectual Capital. *International Review of Management and Marketing*, *14*(6), 274–287. https://doi.org/10.32479/irmm.17257

Aulia, A. M., Safitri, U. N. C., & Hidayati, C. (2025). *Comparative analysis of the financial performance of PT Ultrajaya Milk Industry & Trading Company Tbk and PT Diamond Food Indonesia Tbk period 2019–2023*. *Journal of Advances in Accounting, Economics and Management,* *2*(3), Article 576. https://doi.org/10.47134/aaem.v2i3.576

Babaev, D., Savchenko, M., Sberbank, A., Lab, A., Tuzhilin, Umerenkov, D., & Tuzhilin, A. (2019). *Applying Deep Learning to Credit Loan Applications*. https://doi.org/10.1145/3292500.3330693

Bank Negara Malaysia. (n.d.). *CCRIS Report - Bank Negara Malaysia*. Retrieved July 13, 2025, from https://www.bnm.gov.my/ccris

Bhatt, A. (2022). Document Automation Using Artificial Intelligence. *International Journal for Research in Applied Science and Engineering Technology*, *10*(9), 1365–13169. https://doi.org/10.22214/ijraset.2022.46839

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

CIMB. (2022). *6 Reasons Your Business Loan Application in Malaysia Was Denied.* Business Insights. https://www.cimb.com.my/en/business/business-insights/gain-business-insight/six-reasons-your-business-loan-in-malaysia-was-denied.html

Credit Bureau Malaysia. (n.d.-a). *3 Errors to Watch Out for in Your Credit Report.* https://creditbureau.com.my/3-errors-to-watch-out-for-in-your-credit-report/

Credit Bureau Malaysia. (n.d.-b). *Introducing MySCoRE and MyBizSCoRE*. Retrieved July 13, 2025, from https://creditbureau.com.my/myscore-and-mybizscore/

Credit Bureau Malaysia. (n.d.-c). *What is a credit report?* Retrieved July 13, 2025, from https://creditbureau.com.my/what-is-credit-report/

CTOS Data Systems. (n.d.-a). *Business credit report*. Retrieved July 13, 2025, from https://ctoscredit.com.my/business

CTOS Data Systems. (n.d.-b). *How to read CTOS report (company) – CTOS – Malaysia's leading credit reporting agency*. Retrieved July 13, 2025, from https://ctoscredit.com.my/how-to-read-ctos-report-company/

Department of Statistics Malaysia. (2025). *2024 Micro, small & medium enterprises*. Department of Statistics Malaysia. https://open.dosm.gov.my/publications?page=1

Fišera, B., Horváth, R., & Melecký, M. (2019). *Basel III implementation and SME financing: Evidence for emerging markets and developing economies* (Policy Research Working Paper No. 9069). World Bank. https://espanol.enterprisesurveys.org/content/dam/enterprisesurveys/documents/research-1/SME%20Financing.pdf

Hsu, E., Malagaris, I., Kuo, Y.-F., Sultana, R., & Roberts, K. (2022). Deep learning-based NLP data pipeline for EHR-scanned document information extraction. *JAMIA Open*, *5*(2). https://doi.org/10.1093/jamiaopen/ooac045

Immanuel, J., & Thoyyibah, T. (2025). *Assessing the usability and user acceptance of an e-commerce application through a combined SUMI and TAM framework.* Proceedings of the 5th International Conference on Advances in Electrical, Electronics and Computing Technology (EECT). IEEE. https://doi.org/10.1109/EECT64505.2025.10966960

Jalil, M. F. (2021). Microfinance towards micro-enterprises development in rural Malaysia through digital finance. *Discover Sustainability*, *2*(1). https://doi.org/10.1007/s43621-021-00066-3

Jemai, J., & Zarrad, A. (2023). Feature Selection Engineering for Credit Risk Assessment in Retail Banking. *Information*, *14*(3), 200. https://doi.org/10.3390/info14030200

Katoch, R., & Rani, P. (2023). A Frequency Assessment of Prevalent Prevention Strategies in order to Manage Banks' NPAs in MSME Loans. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(4), 65–80. https://doi.org/10.17762/ijritcc.v11i4.6382

Koç, O., Ugur, O., & Kestel, A. S. (2023). *The Impact of Feature Selection and Transformation on Machine Learning Methods in Determining the Credit Scoring*. ArXiv.org. https://doi.org/10.48550/arXiv.2303.05427

Krishnan, G., & Rani, A. O. (2024). *Market-based Financing for SMEs in Malaysia*. https://www.icmr.my/wp-content/uploads/2024/02/ICMR_SME-Financing-Report_FINAL_23022024.pdf

Kyriazopoulos, G. (2019). *Credit risk evaluation and rating for SMEs using statistical approaches: The case of European SMEs manufacturing sector*. *Journal of Applied Finance & Banking,* *9*(5), 59–83. http://www.scienpress.com/Upload/JAFB/Vol%209_5_4.pdf

Laborda, J., & Ryoo, S. (2021). Feature Selection in a Credit Scoring Model. *Mathematics*, *9*(7), 746. https://doi.org/10.3390/math9070746

Lomas, R., & Reeta. (2024). AI-Driven FinTech Solutions for Financial Inclusion: A Study on MSME Sector Empowerment. *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*, 1887–1891. https://doi.org/10.1109/icac2n63387.2024.10895674

Ma, M.-W., Gao, X.-S., Zhang, Z.-Y., Shang, S.-Y., Jin, L., Liu, P.-L., Feng Lv, Ni, W., Han, Y.-C., & Zong, H. (2023). Extracting laboratory test information from paper-based reports. *BMC Medical Informatics and Decision Making*, *23*(1). https://doi.org/10.1186/s12911-023-02346-6

Mansor, A., Ab Hamid, A. S., Jamal Abdul, J. H., & Muhamad Yusof, N. (2023). Determining the Credit Score Ranking of Malaysian Publicly Traded Companies Using A Merging of The Kmv Merton Model, Financial Ratios and Credit Ratings. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *13*(2). https://doi.org/10.6007/ijarafms/v13-i2/17303

Mohd Hafiz bin Bakar, Hainnuraqma binti Rahim, & Siti Norbaya binti Yahaya. (2022). *Early warning signals on credit risk mitigation among SMEs in Malaysia*. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *12*(3), 750–762. https://doi.org/10.6007/IJARAFMS/v12-i3/14968

National Entrepreneur and SME Development Council. (2025-a). *Economic performance and outlook 2023/24: Advancing productivity and MSMEs' participation in the value chain*. SME Corp. Malaysia. https://www.smeinfo.com.my/msme-insights-2023-24-new-release-2/

National Entrepreneur and SME Development Council. (2025-b). *MSME insights 2023/24: Advancing productivity and MSMEs' participation in the value chain* [Main Report]. SME Corp. Malaysia. https://www.smeinfo.com.my/msme-insights-2023-24-new-release-2/

Nurani, N., Intan Permatasari, R. L., Khalik, A., Hamzah, M., & Nurhani, N. (2025). The Role of Accounting Literacy in Improving the Financial Performance of SMEs: A Study on Micro Entrepreneur Community in Indonesia. *Golden Ratio of Community Services and Dedication*, *5*(2), 28–39. https://doi.org/10.52970/grcsd.v5i2.1451

Osman, F., & Rahman, H. (2024). Implementing AI-Based Credit Risk Models for Small and Medium Enterprises: A Comparative Analysis with Traditional Risk Assessment Approaches. *Algorithms, Computational Theory, Optimization Techniques, and Applications in Research Quarterly*, *14*(7), 1–19. https://ispiacademy.com/index.php/ACORQ/article/view/2024-JULY-04

Ozag, D., & Duguma, B. (2004). *The relationship between cognitive processes and perceived usefulness: An extension of TAM2*. East Carolina University & University of Maryland-University College. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ff424ca05e439b254f7cad5dbff68cf2fbf12750

Pambudi, S. A. L. (2022). How Far Has Our MSMEs Credit Underwriting Assessment in Indonesian Commercial Banks Progressed? *Journal of Business and Management Review*, *3*(10), 717–739. https://doi.org/10.47153/jbmr310.4942022

Rajamani, K., Akbar Jan, N., Subramani, A. K., & Nirmal Raj, A. (2022). Access to Finance: Challenges Faced by Micro, Small, and Medium Enterprises in India. *Engineering Economics*, *33*(1), 73–85. https://doi.org/10.5755/j01.ee.33.1.27998

Rijcken, E., Zervanou, K., Mosteiro, P., Scheepers, F., Spruit, M., & Kaymak, U. (2025). Machine learning vs. rule-based methods for document classification of electronic health records within mental health care—A systematic literature review. *Natural Language Processing Journal*, *10*, 100129. https://doi.org/10.1016/j.nlp.2025.100129

Saygılı, E., Saygılı, A. T., & Isik, G. (2019). *An analysis of factors affecting credit scoring performance in SMEs*. Ege Academic Review, 19(2), 159–171. https://doi.org/10.21121/eab.526093

Shakila Saad, Wan Nurhadani Wan Jaafar, Siti Jasmida Jamil, & Nooraihan Abdullah. (2025). *Developing a credit scoring of the SMEs manufacturing based on multi criteria decision making (MCDM) algorithm*. *Applied Mathematics and Computational Intelligence (AMCI)*, *14*(1), 12–36. https://doi.org/10.58915/amci.v14i1.195

Sharil Abdul Rahman. (2024). *High Court rules CTOS cannot create credit scores; orders to pay RM200,000 to resort owner*. Malay Mail. https://www.malaymail.com/news/malaysia/2024/03/11/high-court-rules-ctos-cannot-create-credit-scores-orders-to-pay-rm200000-to-resort-owner/122857

Shreya, & Pathak, H. (2025). *Explainable Artificial Intelligence Credit Risk Assessment using Machine Learning*. ArXiv.org. https://arxiv.org/abs/2506.19383

Stevenson, M. P. (2024). Novel applications of advanced predictive analytics and artificial intelligence to improve SME competitiveness and access to funding - ePrints Soton. *Soton.ac.uk*. https://eprints.soton.ac.uk/492362/1/AdvancedAnalyticsAIforSME_Thesis-6.pdf

Vendy, V., & Sucahyati, D. (2022). *Issues and challenges of adoption of IFRS for SMEs in Malaysia*. *Nusantara Science and Technology Proceedings*, 6–12. https://doi.org/10.11594/nstp.2022.2302

Wang, Z., & Liang, J. (2024). Comparative Analysis of Interpretability Techniques for Feature Importance in Credit Risk Assessment. *Spectrum of Research*, *4*(2). http://spectrumofresearch.com/index.php/sr/article/view/21

Wasiuzzaman, S., Nurdin, N., Abdullah, A. H., & Vinayan, G. (2019). Creditworthiness and access to finance: a study of SMEs in the Malaysian manufacturing industry. *Management Research Review*, *43*(3), 293–310. https://doi.org/10.1108/mrr-05-2019-0221

Wibowo, P., & Sanjaya, S. A. (2024). Predicting Potential Non-Performing Loans Collectibility in MSMEs using Ensemble Stacking with SHAP-Boosting Algorithm. *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 479–484. https://doi.org/10.1109/isriti64779.2024.10963398

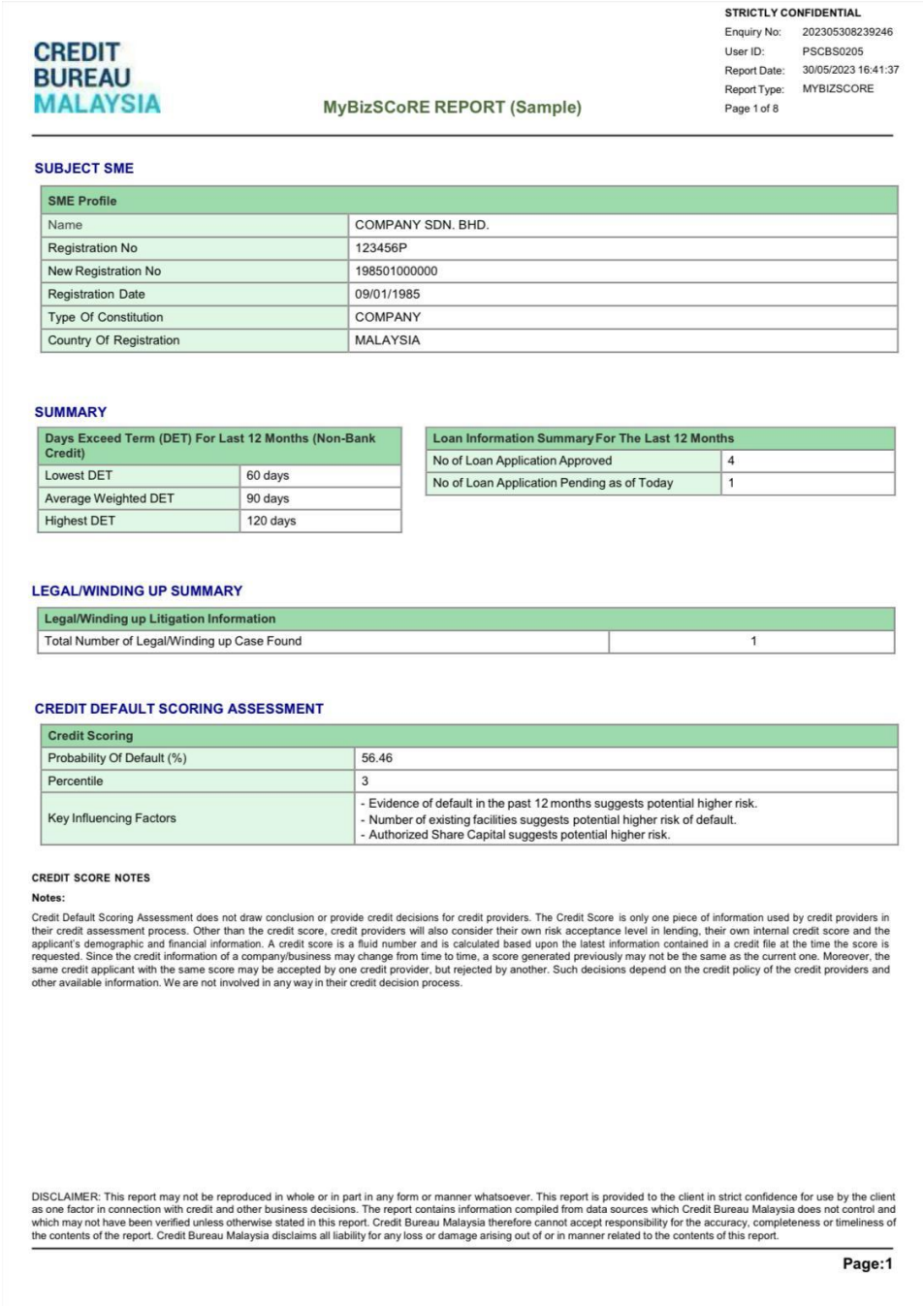
Youssef, K. Y. (2025). *Evaluating the performance of non-profit organizations using trend analysis: The future impacts of the present performance*. *Arab Journal of Administration, 45*(2), 227–238. https://doi.org/10.21608/aja.2022.131632.1229

Yuan, Y. (2020). A CRITICAL REVIEW OF FINANCIAL ACCESSIBILITY AMONG SMALL AND MEDIUM ENTERPRISES IN MALAYSIA. *Journal of International Business, Innovation and Strategic Management*, *4*(1), 66–82. https://www.jibism.org/core_files/index.php/JIBISM/article/view/121

Appendix

Appendix A: Sample Credit Risk Report (MyBizSCoRE)

Figure A.1, A.2 and A.3 shows an example credit risk report used in this study, demonstrating the layout and financial data presentation typically found in CBM-issued PDFs.



STRICTLY CONFIDENTIAL
Enquiry No: 202305308239246
User ID: PSCBS0205
Report Date: 30/05/2023 16:41:37
Report Type: MYBIZSCORE
Page 1 of 8

CREDIT BUREAU MALAYSIA

MyBizSCoRE REPORT (Sample)

SUBJECT SME

SME Profile	
Name	COMPANY SDN. BHD.
Registration No	123456P
New Registration No	198501000000
Registration Date	09/01/1985
Type Of Constitution	COMPANY
Country Of Registration	MALAYSIA

SUMMARY

Days Exceed Term (DET) For Last 12 Months (Non-Bank Credit)	
Lowest DET	60 days
Average Weighted DET	90 days
Highest DET	120 days

Loan Information Summary For The Last 12 Months	
No of Loan Application Approved	4
No of Loan Application Pending as of Today	1

LEGAL/WINDING UP SUMMARY

Legal/Winding up Litigation Information	
Total Number of Legal/Winding up Case Found	1

CREDIT DEFAULT SCORING ASSESSMENT

Credit Scoring	
Probability Of Default (%)	56.46
Percentile	3
Key Influencing Factors	- Evidence of default in the past 12 months suggests potential higher risk. - Number of existing facilities suggests potential higher risk of default. - Authorized Share Capital suggests potential higher risk.

CREDIT SCORE NOTES

Notes:

Credit Default Scoring Assessment does not draw conclusion or provide credit decisions for credit providers. The Credit Score is only one piece of information used by credit providers in their credit assessment process. Other than the credit score, credit providers will also consider their own risk acceptance level in lending, their own internal credit score and the applicant's demographic and financial information. A credit score is a fluid number and is calculated based upon the latest information contained in a credit file at the time the score is requested. Since the credit information of a company/business may change from time to time, a score generated previously may not be the same as the current one. Moreover, the same credit applicant with the same score may be accepted by one credit provider, but rejected by another. Such decisions depend on the credit policy of the credit providers and other available information. We are not involved in any way in their credit decision process.

DISCLAIMER: This report may not be reproduced in whole or in part in any form or manner whatsoever. This report is provided to the client in strict confidence for use by the client as one factor in connection with credit and other business decisions. The report contains information compiled from data sources which Credit Bureau Malaysia does not control and which may not have been verified unless otherwise stated in this report. Credit Bureau Malaysia therefore cannot accept responsibility for the accuracy, completeness or timeliness of the contents of the report. Credit Bureau Malaysia disclaims all liability for any loss or damage arising out of or in manner related to the contents of this report.

Page:1

Figure A.1: Report Summary Section (MyBizSCoRE Sample)

FINANCIAL STATEMENT*

Summary of Financial Information					
Auditor	T.H.YEW & CO	YW SOO & CO	KP	KP	-
Auditor Address	NO. 1/2 JALAN CHEW BOONJUAN 30250 IPOH	NO 123 JALAN LEONG SIN NAM 30300 IPOH	123A JALAN SS A/B 47300 PETALING JAYA	123A JALAN SS A/B 47300 PETALING JAYA	-
Exempt Private Company	-	-	-	-	-
Financial year end	31/05/2021	31/05/2020	31/05/2019	31/05/2018	-
Unqualified reports (Y/N)	Y	Y	Y	Y	-
Consolidated accounts (Y/N)	N	N	N	N	-
Date of tabling	30/11/2021	30/11/2020	30/11/2019	30/11/2018	-

Balance Sheet Items					
Financial year end	31/05/2021	31/05/2020	31/05/2019	31/05/2018	-
Non-current assets	2,780,528.00	2,873,948.00	2,869,109.00	1,408,482.00	-
Current assets	5,379,602.00	4,317,295.00	4,993,508.00	5,151,592.00	-
Non-current liabilities	2,319,009.00	2,182,047.00	754,430.00	655,275.00	-
Current liabilities	4,362,254.00	3,556,732.00	5,553,909.00	4,416,194.00	-
Share capital	750,000.00	750,000.00	750,000.00	750,000.00	-
Reserves	0.00	0.00	0.00	0.00	-
Retained earning	728,867.00	702,464.00	804,278.00	738,605.00	-
Minority interests	0.00	0.00	0.00	0.00	-

Income Statement Items					
Financial year end	31/05/2021	31/05/2020	31/05/2019	31/05/2018	-
Revenue	16,750,468.00	13,444,606.00	16,322,540.00	16,280,553.00	-
Profit / (loss) before tax	115,556.00	-101,669.00	115,071.00	63,483.00	-
Profit / (loss) after tax	91,859.00	-101,814.00	65,673.00	33,691.00	-
Net dividend	0.00	0.00	0.00	0.00	-
Minority Interests	0.00	0.00	0.00	0.00	-

Financial Ratios					
Financial year end	31/05/2021	31/05/2020	31/05/2019	31/05/2018	-
Current Ratio (Times)	0.8	0.5	0.4	1.1	-
Gearing Ratio (Times)	0.8	2.5	1.9	0.6	-
ROCE (Return on Capital Employed)%	2.0	-2.7	-36.7	-2.3	-
Assets Turnover Ratio (Times)	0.5	0.4	0.3	0.3	-
Earnings Per Share (RM p/share)	0.0	-0.1	-0.7	-0.1	-

Figure A.2: Financial Information and Ratio

Loan Information

No	Date/ R&R Date	ST	Capacity	Lender Type	Facility/ App Type	Total Outstanding Balance (RM)	Date Balance Updated	Limit / Instl Amt (RM)	Prin. Repmt. Term	Col Type	Conduct Of Account For Last 12 Months												LGL STS	Date Status Updated
											2023											2022		
Outstanding Credit											M	A	M	F	J	D	N	O	S	A	J	J		
1	11/01/2021		Own	Bank A				50,000.00																
		O			Other term loan/Finance (include personal loan/ finance)	51,267.00	30/04/2023	910.00	MTH	CLEAN		11	9	8	7	6	5	4	3	2	1	2		
2	10/09/2021		Own	Bank B				100,000.00		FIN GUARANTEE														
		O			Other term loan/Finance (include personal loan/ finance)	103,860.00	30/04/2023	1,595.00	MTH			5	4	3	2	1	0	0	0	0	0	0		
3	31/12/2021		Own	Bank C				500,000.00															Jdgmtord/ ordsal	17/03/2023
		O			Other term loan/Finance (include personal loan/ finance)	541,460.00	30/04/2023	8,483.00	MTH	FIN GUARANTEE		11	10	9	8	7	6	0	0	0	0	0		
					Total Outstanding Balance:	696,587.00	Total Limit:	650,000.00																
Special Attention Account																								
1	22/01/2020		Own	Bank D	Purchase of passenger cars		30/04/2023																Jdgmtord/ ordsal	01/02/2023
2	15/01/2021		Own	Bank E	Other term loan/Finance (include personal loan/ finance)		30/04/2023																Jdgmtord/ ordsal	02/12/2022
Credit Application																								
1	25/07/2022	A	Joint	Bank F	N			50,000.00																
					FNINSOFN																			
2	11/01/2023	P	Own	Bank G	N			500,000.00																
					OTLNFNCE																			

Figure A.3: Loan Performance and Repayment Records