Name: Claudine Linda Wa Nciko

Student Number: 169375

Lecturer: Dr. John Olukuru

Assistant Instructor: Joan Ngugi

BIG DATA IN FINANCE AND BANKING: EXAMINATION: DSA 8504

QUESTION 2

1. Strategic Data Splitting(3mks) Jenga Microfinance current analysts split data randomly for training/testing. Explain why this is dangerous in a lending environment. What is the best splitting strategy based on disbursement_date, and explain how it helps improve model reliability.

R/

Random data splitting in a lending environment is risky because it ignores the temporal structure inherent in loan data. This can lead to look-ahead bias, where the model inadvertently learns from future disbursement and repayment behaviors that would not be available at the time of making real-world credit decisions. Such leakage inflates performance metrics and results in unreliable models that may fail in production.

The best splitting strategy in this context is time-based splitting using the disbursement_date. This approach involves sorting loans chronologically and assigning the earliest 70–80% of data for training, the next 10–15% for validation, and the most recent 10–15% for testing. This method mirrors real-world deployment, where models predict future loan performance based solely on historical data. It enhances reliability by exposing the model to genuine temporal shifts in customer behavior and macroeconomic conditions, ensuring robustness and generalizability in dynamic lending environments.

2. Variable Treatment & Governance (8 marks) Jenga Microfinance collects over 15 features, including sensitive fields like marital status and education. a) Propose a logic for: Selecting variables using WOE/IV Handling categorical variables like employment_status or residence_type Binning continuous variables for modelling b) List two features that you would exclude from the model and justify why — either from a regulatory, fairness, or predictive standpoint.

R/

(a) A structured approach is essential for variable treatment and governance. Variable selection begins with calculating Weight of Evidence (WOE) and Information Value (IV). WOE highlights the strength and direction of a variable's relationship with default risk, while IV helps assess its predictive power. Variables with IV above 0.1 are typically retained, provided they show stable and monotonic WOE patterns over time.

For categorical variables such as *employment_status* or *residence_type*, WOE encoding is ideal as it converts categories into numerical values that reflect their relative risk. Rare categories should be grouped to improve stability, and the final groupings must align with business logic.

Continuous variables like income or loan amount are binned to capture non-linear risk relationships. The process involves initial binning, followed by merging adjacent bins with similar WOE, ensuring stability and interpretability. Finer bins may be used at the tails to capture outliers, and final bins should be validated with business experts.

(b) Two features I would exclude are:

- **Marital status** due to fairness and regulatory concerns, as it may introduce unintended bias without significantly improving predictive power.

- **Education** if it shows low predictive value or inconsistent patterns over time, it should be excluded to maintain model robustness and avoid reinforcing social inequalities.

3. Cutoff Strategy Design (5 marks) You've built a scorecard model using logistic regression. Now the business wants to deploy it to auto-approve loans. a) You observe that: 70% of non-defaulters score above 650.80% of defaulters score below 580. Which cutoff strategy would you recommend for Automatic approval, Manual review and Rejection of loan applications based on the scores: b) What two types of validation would you run to monitor model performance over time once deployed?

R/

(a) Based on the score distribution, a three-tier cutoff strategy is appropriate:

- **Automatic Approval:** Approve applicants scoring **above 650**, as 70% of non-defaulters fall in this range, indicating low risk.
- **Manual Review:** Applicants scoring between **580 and 650** should undergo manual review, as this range includes overlapping risk segments that require further scrutiny.
- **Rejection:** Reject applicants scoring **below 580**, since 80% of defaulters are concentrated here, signaling high credit risk.

This strategy aligns approval decisions with risk profiles, balancing business growth and portfolio quality.

(b) To monitor model performance post-deployment:

1. **Population Stability Index (PSI):** Used to detect shifts in score distribution over time. A PSI above 0.1 indicates a shift worth monitoring, while values above 0.2 suggest the model may require retraining.

2. **Characteristic Stability Analysis:** Tracks changes in the predictive power (WOE/IV) of individual features. Declines in IV or unstable WOE trends signal feature degradation, enabling proactive updates without full model rebuilds.

These validations ensure long-term model reliability and regulatory compliance.

4. Bias, Drift & Ethical Traps (4 marks) Jenga Microfinance is committed to responsible AI. You're asked to audit your scorecard. a) Identify one potential source of data leakage or bias that could affect fairness. b) Explain how model drift might appear in this lending context — and how to detect/respond to it.

R/

(a) A key source of bias in credit scoring is the inclusion of variables that act as proxies for sensitive attributes. For example, *residence type* may unintentionally reflect socioeconomic status, disadvantaging applicants from lower-income backgrounds. Such features can lead to discriminatory outcomes and reinforce systemic inequalities. To mitigate this, proxy variables should be reviewed and, where necessary, replaced with more objective alternatives such as *length of residence* or *housing expense ratio* that capture risk without introducing unfair bias.

(b) Model drift may occur as economic conditions or applicant characteristics evolve such as rising unemployment, inflation, or shifts in borrower behavior. This drift causes the model's predictive accuracy to deteriorate over time. It can be detected through ongoing monitoring using tools like the Population Stability Index (PSI), characteristic stability analysis, and tracking prediction errors across time segments. To respond, Jenga should implement regular model reviews, retrain with recent data, recalibrate thresholds, and consider adaptive strategies like including macroeconomic indicators to maintain fairness and performance.

QUESTION 3

1. Score Sensitivity Design (3 marks) Jenga Microfinance leadership wants a scorecard where small differences in customer risk are clearly reflected in score differences. Should the team choose a high PDO (e.g., 70) or a low PDO (e.g., 30)? Explain how this choice affects score sensitivity and interpretation.

R/

To reflect small differences in customer risk more clearly, Jenga Microfinance should choose a higher PDO (e.g., 70). The PDO controls how many score points are needed to double the odds of non-default. A higher PDO increases score granularity, meaning that even small changes in the predicted probability of default result in noticeable score differences.

This enhances score sensitivity, allowing the model to better distinguish between applicants with similar risk levels. It also reduces the chance that minor model fluctuations cause applicants to cross key decision thresholds. As a result, decision-making becomes more stable, and policy cutoffs can be set with greater precision supporting fairer, more strategic credit allocation.

2. Offset Alignment (3 marks) The Head of Risk insists that customers with odds of 20:1 (good:bad) should receive a score of exactly 600. What is the purpose of Offset in this case? What would happen if Offset were incorrectly set too high or too low?

R/

The **Offset** aligns the score scale with business-defined baselines. In this case, it ensures that customers with **20:1 odds (good:bad)** receive a score of **600**, anchoring the scorecard at a meaningful reference point. This makes the score interpretable and consistent with internal policies.

If the Offset is set **too high**, all scores shift upward, making risky applicants appear more creditworthy and increasing the risk of over-approvals. If set **too low**, even low-risk applicants may fall below approval thresholds, leading to unnecessary rejections and reduced loan volumes.

Correct calibration of the Offset is essential to maintain alignment between credit scores, risk levels, and approval policies ensuring both fairness and financial performance.

3. Score Meaning & Customer Profiles (4 marks) Two customers apply for a loan: Customer A: $p = 0.4$ Customer B: $p = 0.1$ a. Without computing, who will receive a higher score and why? b. What does a higher score represent in terms of default risk and odds?

R/

(a) Customer **B** will receive a higher score than Customer **A**. This is because a lower predicted probability of default (p = 0.1) translates into better odds of repayment. The scoring formula is designed such that higher odds (i.e., lower risk) produce higher scores. Since Customer B is less likely to default, the model assigns them a higher score, reflecting stronger creditworthiness.

(b) A **higher score** represents **lower default risk** and **higher odds of repayment**. The score increases logarithmically with the odds of non-default, meaning every set number of points (defined by the PDO) corresponds to a doubling of the odds. Higher scores indicate customers who are statistically more reliable, making them better candidates for automatic approvals, higher credit limits, and preferential pricing supporting sound risk-based decision-making.

4. Cutoff & Policy Trade-offs (5 marks) Jenga Microfinance is considering the following score thresholds: Score > 650 → Auto-approve Score 580–650 → Manual review Score < 580 → Reject a. What business risk does Jenga Microfinance reduce by auto-approving only customers above 650? b. What trade-offs might arise from putting too many customers into the "manual review" range? c. How could the wrong choice of PDO or Offset misclassify customers across these thresholds?

R/

(a) By auto-approving only customers with scores above **650**, Jenga Microfinance reduces **credit risk** by limiting automatic approvals to highly creditworthy applicants. This protects the institution from approving high-risk loans without human oversight and also mitigates **model risk**, as only low-risk profiles bypass manual review. It ensures automated decisions are conservative, reducing the likelihood of default-driven losses.

(b) A broad manual review range (580–650) creates several trade-offs. It increases **operational costs** and slows down the approval process due to the need for human intervention. This can lead to **delays**, **inconsistent decisions**, and **customer dissatisfaction**, especially if applicants abandon the process. It also limits **scalability**, as the organization may struggle to keep up with application volumes without significantly expanding its credit review team.

(c) The **PDO** affects score sensitivity. If set too low, scores will bunch together, reducing the model's ability to distinguish between risk levels—causing some applicants to be wrongly approved or rejected. If set too high, small differences in risk are exaggerated, pushing too many applicants into manual review or the wrong decision band. An incorrect **Offset** shifts the entire score distribution. A high Offset may inflate scores, approving risky applicants, while a low Offset may suppress scores, rejecting creditworthy ones. Both misconfigurations lead to **misclassification**, harming both risk control and customer experience.