Lesson 2.2 Quiz: Credit Scoring — From FICO to ML

Module 2: The Access Problem

Prof. Dr. Joerg Osterrieder

Digital Finance — BSc Course

Q1: Purpose of a Credit Score

What is the primary purpose of a credit score?

- (A) To determine the exact amount a borrower should receive
- (B) To predict the probability that a borrower will default within a defined time horizon
- (C) To calculate the interest rate on a mortgage
- (D) To verify the borrower's identity

Q1: Purpose of a Credit Score

What is the primary purpose of a credit score?

- (A) To determine the exact amount a borrower should receive
- (B) To predict the probability that a borrower will default within a defined time horizon
- (C) To calculate the interest rate on a mortgage
- (D) To verify the borrower's identity

Answer: (B) A credit score is fundamentally a prediction of the probability of default (PD), typically within 12 months.

Q2: FICO Score Components

Which factor carries the **highest weight** in a traditional FICO credit score?

A. Length of credit history (15%)
B. Amounts owed / credit utilisation (30%)
C. Payment history (35%)
D. Credit mix (10%)

Q2: FICO Score Components

Which factor carries the **highest weight** in a traditional FICO credit score?

- (A) Length of credit history (15%)
- (B) Amounts owed / credit utilisation (30%)
- (C) Payment history (35%)
- (D) Credit mix (10%)

Answer: (C) Payment history (on-time vs. late payments) accounts for 35% of a FICO score—the single largest component.

Q3: WoE Interpretation

A bin in a credit scorecard has a Weight of Evidence (WoE) of $+0.40$. What does this indicate?

- (A) The bin contains exactly 40% of all defaults
- (B) The bin has proportionally more goods (non-defaulters) than bads (defaulters)
- (C) The bin has proportionally more bads than goods
- (D) The feature is not predictive

Q3: WoE Interpretation

A bin in a credit scorecard has a Weight of Evidence (WoE) of $+0.40$. What does this indicate?

- (A) The bin contains exactly 40% of all defaults
- (B) The bin has proportionally more goods (non-defaulters) than bads (defaulters)
- (C) The bin has proportionally more bads than goods
- (D) The feature is not predictive

Answer: (B) A positive WoE means the distribution of goods in this bin exceeds the distribution of bads—the bin is associated with lower default risk.

Q4: Information Value Thresholds

An analyst calculates the Information Value (IV) of a candidate feature and obtains IV $= 0.03$. How should this feature be classified?

- **A** Strong predictor
- **B** Medium predictor
- **C** Weak predictor
- **D** Suspicious — check for data leakage

Q4: Information Value Thresholds

An analyst calculates the Information Value (IV) of a candidate feature and obtains IV $= 0.03$. How should this feature be classified?

- **A** Strong predictor
- **B** Medium predictor
- **C** Weak predictor
- **D** Suspicious — check for data leakage

Answer: (C) IV in the range 0.02–0.10 indicates a weak predictor. It may still contribute in a model but offers limited separation power on its own.

Q5: WoE Calculation

A bin contains 20% of all goods and 10% of all bads. What is the WoE of this bin?

- (A) $\ln(0.10/0.20) = -0.693$
- (B) $\ln(0.20/0.10) = 0.693$
- (C) $0.20 - 0.10 = 0.10$
- (D) $0.20 \times 0.10 = 0.02$

Q5: WoE Calculation

A bin contains 20% of all goods and 10% of all bads. What is the WoE of this bin?

- (A) $\ln(0.10/0.20) = -0.693$
- (B) $\ln(0.20/0.10) = 0.693$
- (C) $0.20 - 0.10 = 0.10$
- (D) $0.20 \times 0.10 = 0.02$

Answer: (B) WoE $= \ln(\text{Distr. Goods}/\text{Distr. Bads}) = \ln(0.20/0.10) = \ln(2) = 0.693$.

Q6: Expected Loss Calculation

A personal loan has PD = 4%, LGD = 45%, and EAD = \$20,000. What is the expected loss?

A. \$36
B. \$360
C. \$900
D. \$3,600

Q6: Expected Loss Calculation

A personal loan has PD = 4%, LGD = 45%, and EAD = \$20,000. What is the expected loss?

- (A) \$36
- (B) \$360
- (C) \$900
- (D) \$3,600

Answer: (B) EL = PD $\times$ LGD $\times$ EAD = $0.04 \times 0.45 \times \$20{,}000 = \$360$.

Q7: Gini from AUC

A credit scoring model achieves an AUC of 0.78 on the validation dataset. What is the corresponding Gini coefficient?

- A 0.39
- B 0.56
- C 0.78
- D 1.56

Q7: Gini from AUC

A credit scoring model achieves an AUC of 0.78 on the validation dataset. What is the corresponding Gini coefficient?

- (A) 0.39
- (B) 0.56
- (C) 0.78
- (D) 1.56

Answer: (B) $\text{Gini} = 2 \times \text{AUC} - 1 = 2 \times 0.78 - 1 = 0.56$.

Q8: ROC Curve Interpretation

A credit model has AUC = 0.50 on out-of-time validation data. What does this mean?

A. The model is perfectly calibrated
B. The model has no discriminatory power—equivalent to random guessing
C. The model has moderate discrimination
D. The model is overfitting

Q8: ROC Curve Interpretation

A credit model has AUC = 0.50 on out-of-time validation data. What does this mean?

- (A) The model is perfectly calibrated
- (B) The model has no discriminatory power—equivalent to random guessing
- (C) The model has moderate discrimination
- (D) The model is overfitting

Answer: (B) AUC = 0.50 means the model cannot distinguish goods from bads—it performs no better than random ordering.

Q9: PSI Interpretation

A quarterly monitoring report shows PSI $= 0.18$ for the application scorecard. What action is recommended?

- **A** No action — this is within acceptable limits
- **B** Investigate the source of the shift
- **C** Immediately rebuild the model
- **D** Retire the scorecard

Q9: PSI Interpretation

A quarterly monitoring report shows PSI = 0.18 for the application scorecard. What action is recommended?

- (A) No action — this is within acceptable limits
- (B) Investigate the source of the shift
- (C) Immediately rebuild the model
- (D) Retire the scorecard

Answer: (B) PSI between 0.10 and 0.25 indicates moderate population shift. The model should be investigated to determine if recalibration or rebuild is needed.

Q10: IV Calculation

A feature has two bins. Bin 1: Distr. Goods $= 0.60$, Distr. Bads $= 0.40$, WoE $= 0.405$. Bin 2: Distr. Goods $= 0.40$, Distr. Bads $= 0.60$, WoE $= -0.405$. What is the IV?

- A 0.000
- B 0.081
- C 0.162
- D 0.405

Q10: IV Calculation

A feature has two bins. Bin 1: Distr. Goods $= 0.60$, Distr. Bads $= 0.40$, WoE $= 0.405$. Bin 2: Distr. Goods $= 0.40$, Distr. Bads $= 0.60$, WoE $= -0.405$. What is the IV?

- (A) 0.000
- (B) 0.081
- (C) 0.162
- (D) 0.405

Answer: (C) IV $= (0.60 - 0.40) \times 0.405 + (0.40 - 0.60) \times (-0.405) = 0.081 + 0.081 = 0.162$. This is a medium predictor.

Q11: Score-to-PD Conversion

A scorecard uses the formula $\text{Score} = 600 - 40 \times \ln(\text{PD}/(1-\text{PD}))$. A borrower receives a score of 640. Which statement is correct about their PD?

- **A** PD is higher than the borrower with score 600
- **B** PD is lower than the borrower with score 600
- **C** PD is exactly 50%
- **D** The score is invalid because it exceeds 600

Q11: Score-to-PD Conversion

A scorecard uses the formula Score $= 600 - 40 \times \ln(\text{PD}/(1-\text{PD}))$. A borrower receives a score of 640. Which statement is correct about their PD?

- (A) PD is higher than the borrower with score 600
- (B) PD is lower than the borrower with score 600
- (C) PD is exactly 50%
- (D) The score is invalid because it exceeds 600

Answer: (B) A higher score of 640 means $\ln(\text{PD}/(1-\text{PD}))$ must be more negative, implying lower PD. Higher score = lower default probability.

Q12: Gradient Boosting Hyperparameters

When building a credit scoring model with gradient boosting, a data scientist sets `max_depth = 3` and `learning_rate = 0.05`. What is the rationale?

- **A** Deep trees with fast learning maximise training speed
- **B** Shallow trees with slow learning reduce overfitting and improve generalisation
- **C** Shallow trees cannot capture non-linear relationships
- **D** These settings are only valid for random forests, not gradient boosting

Q12: Gradient Boosting Hyperparameters

When building a credit scoring model with gradient boosting, a data scientist sets `max_depth = 3` and `learning_rate = 0.05`. What is the rationale?

- (A) Deep trees with fast learning maximise training speed
- (B) Shallow trees with slow learning reduce overfitting and improve generalisation
- (C) Shallow trees cannot capture non-linear relationships
- (D) These settings are only valid for random forests, not gradient boosting

Answer: (B) Shallow trees (low `max_depth`) prevent individual trees from memorising the data, and low `learning_rate` requires more trees but produces better generalisation.

Q13: Discrimination vs. Calibration

A credit model has AUC $= 0.82$ but consistently predicts PDs that are 3 percentage points lower than observed default rates. What is the diagnosis?

- **A** The model has poor discrimination and poor calibration
- **B** The model has good discrimination but poor calibration
- **C** The model has poor discrimination but good calibration
- **D** The model has good discrimination and good calibration

Q13: Discrimination vs. Calibration

A credit model has AUC = 0.82 but consistently predicts PDs that are 3 percentage points lower than observed default rates. What is the diagnosis?

- **A** The model has poor discrimination and poor calibration
- **B** The model has good discrimination but poor calibration
- **C** The model has poor discrimination but good calibration
- **D** The model has good discrimination and good calibration

Answer: (B) AUC = 0.82 shows good ranking ability (discrimination), but systematic underestimation of PDs indicates poor calibration. The model ranks correctly but the probabilities are wrong.

Q14: Alternative Data Trade-off

A FinTech lender in East Africa uses mobile airtime top-up patterns to score thin-file borrowers. Which risk is **most specific** to this alternative data source?

- A The logistic regression model may overfit
- B Mobile usage patterns may change rapidly, causing model drift
- C The FICO score range is too narrow for this population
- D The model's AUC will always be below 0.50

Q14: Alternative Data Trade-off

A FinTech lender in East Africa uses mobile airtime top-up patterns to score thin-file borrowers. Which risk is **most specific** to this alternative data source?

- (A) The logistic regression model may overfit
- (B) Mobile usage patterns may change rapidly, causing model drift
- (C) The FICO score range is too narrow for this population
- (D) The model's AUC will always be below 0.50

Answer: (B) Mobile usage patterns are volatile—new apps, pricing plan changes, and seasonal behaviours can rapidly shift the data distribution, causing model drift (high PSI).

Q15: Proxy Discrimination

A credit model uses zip code as a feature. The zip code has high IV and improves AUC. However, zip codes are highly correlated with race in the US. What is the issue?

- **A** The model will have low AUC because zip code is not predictive
- **B** Zip code acts as a proxy for a protected characteristic, creating illegal discrimination
- **C** Zip codes should always be included because they improve accuracy
- **D** This is only a problem if race is directly included in the model

Q15: Proxy Discrimination

A credit model uses zip code as a feature. The zip code has high IV and improves AUC. However, zip codes are highly correlated with race in the US. What is the issue?

- (A) The model will have low AUC because zip code is not predictive
- (B) Zip code acts as a proxy for a protected characteristic, creating illegal discrimination
- (C) Zip codes should always be included because they improve accuracy
- (D) This is only a problem if race is directly included in the model

Answer: (B) Even without using race directly, zip code can serve as a proxy, leading to disparate impact. Under ECOA and fair lending laws, this constitutes illegal discrimination if the feature's benefit does not outweigh its discriminatory effect.

Q16: Feedback Loop in Credit Scoring

A bank declines 30% of applicants based on its scorecard. Declined applicants never generate repayment data. Over time, what happens to the model?

- **A** The model improves because it only trains on successful loans
- **B** The model becomes biased toward the approved population, potentially reinforcing initial biases
- **C** Nothing changes because the model was validated at deployment
- **D** The model automatically corrects for missing data

Q16: Feedback Loop in Credit Scoring

A bank declines 30% of applicants based on its scorecard. Declined applicants never generate repayment data. Over time, what happens to the model?

- (A) The model improves because it only trains on successful loans
- (B) The model becomes biased toward the approved population, potentially reinforcing initial biases
- (C) Nothing changes because the model was validated at deployment
- (D) The model automatically corrects for missing data

Answer: (B) This is a feedback loop: the model never observes outcomes for declined applicants, so it cannot learn that some would have repaid. This reinforces the initial deny pattern and can perpetuate historical biases.

Q17: ML vs. Logistic Regression

Compared to a well-built logistic regression scorecard, gradient boosting typically shows the largest AUC improvement in which scenario?

- A When only 5 traditional bureau features are available
- B When hundreds of features including alternative data and interaction effects are available
- C When the training dataset has fewer than 100 observations
- D When all features have linear relationships with log-odds of default

Q17: ML vs. Logistic Regression

Compared to a well-built logistic regression scorecard, gradient boosting typically shows the largest AUC improvement in which scenario?

- (A) When only 5 traditional bureau features are available
- (B) When hundreds of features including alternative data and interaction effects are available
- (C) When the training dataset has fewer than 100 observations
- (D) When all features have linear relationships with log-odds of default

Answer: (B) Gradient boosting excels when many features with non-linear effects and interactions are available. With few features and linear relationships, logistic regression performs comparably.

Q18: Model Validation Dimension

A bank's credit model shows AUC = 0.75 at deployment but AUC = 0.62 twelve months later. Which validation dimension has failed?

- A Calibration
- B Concentration
- C Stability (temporal performance)
- D Stress testing

Q18: Model Validation Dimension

A bank's credit model shows AUC = 0.75 at deployment but AUC = 0.62 twelve months later. Which validation dimension has failed?

- (A) Calibration
- (B) Concentration
- (C) Stability (temporal performance)
- (D) Stress testing

Answer: (C) A 13-point AUC decline over 12 months indicates temporal instability—the model's discriminatory power has degraded, likely due to population drift or macroeconomic change.

Q19: Two-Model Approach Evaluation

A FinTech lender uses a gradient boosting model (AUC = 0.84) for decisioning and a logistic regression surrogate (AUC = 0.76) for generating adverse action reasons. A regulator questions this approach. Which criticism is **most valid**?

- **A** The logistic model is too inaccurate to use at all
- **B** The surrogate may give reasons that do not reflect the actual decision logic of the ML model
- **C** Two models always cost more than one, so this is wasteful
- **D** The gradient boosting model should be abandoned in favour of the logistic model

Q19: Two-Model Approach Evaluation

A FinTech lender uses a gradient boosting model (AUC = 0.84) for decisioning and a logistic regression surrogate (AUC = 0.76) for generating adverse action reasons. A regulator questions this approach. Which criticism is **most valid**?

- (A) The logistic model is too inaccurate to use at all
- (B) The surrogate may give reasons that do not reflect the actual decision logic of the ML model
- (C) Two models always cost more than one, so this is wasteful
- (D) The gradient boosting model should be abandoned in favour of the logistic model

Answer: (B) The core risk is that the surrogate explains a *different* model. The reasons it provides may not accurately reflect why the ML model declined a specific applicant, undermining the legal purpose of adverse action notices.

Q20: Policy Recommendation

A developing country's central bank is deciding whether to allow FinTech lenders to use social media data for credit scoring. Which recommendation best balances inclusion and fairness?

- A Ban all alternative data to protect consumer privacy
- B Allow unrestricted use of social media data to maximise financial inclusion
- C Permit social media data with mandatory fairness testing, opt-in consent, and regular bias audits
- D Require all lenders to use only FICO-equivalent bureau scores

Q20: Policy Recommendation

A developing country's central bank is deciding whether to allow FinTech lenders to use social media data for credit scoring. Which recommendation best balances inclusion and fairness?

- (A) Ban all alternative data to protect consumer privacy
- (B) Allow unrestricted use of social media data to maximise financial inclusion
- (C) Permit social media data with mandatory fairness testing, opt-in consent, and regular bias audits
- (D) Require all lenders to use only FICO-equivalent bureau scores

Answer: (C) A balanced approach allows innovation for inclusion while requiring safeguards: informed consent ensures privacy, fairness testing prevents discrimination, and bias audits catch emerging issues. Blanket bans or unrestricted use both fail to balance competing objectives.