Module 4 Assignment: Analyze Machine Learning and Finances Scenario

Opeyemi Morakinyo

Instructor: Damian Dunbar

Scenario 2. Bias Mitigation in Loan Approval AI

Scenario:
You are part of a team responsible for an AI-driven loan approval system at a bank. The system has been found to show bias against certain demographic groups, leading to discriminatory outcomes. The bank intends to address this bias issue and ensure fair lending practices.

Questions:

a. Explain the concept of algorithmic bias in the context of the loan approval AI system. How does biased decision-making impact the bank's reputation and customer relationships?
b. Outline the steps involved in collecting and curating a diverse dataset to mitigate bias in the loan approval AI model.
c. How might retraining the AI model with a more balanced dataset help reduce bias and discrimination in loan approvals? What challenges might arise during this retraining process?
d. Apart from dataset adjustments, suggest two additional strategies the bank could implement to minimize bias in the loan approval AI system.

Answer:

a. Inadequate or faulty data can lead to biased or incorrect decisions being made by AI systems. Two ways exist for low-quality data to lead to discriminatory or biased decision-making. Insufficiently trained machine learning algorithms may yield imprecise outcomes even when supplied with high-quality data (OECD, 2021). Predisposition of the loan approval algorithm When a lending decision is made unfairly and systemically based on a borrower's age, gender, or ethnicity, it is referred to as an AI system. The bank's reputation and client relationships may suffer from biased decision-making in several ways:

1. Reputation: The bank's reputation may suffer from discriminatory outcomes, which could result in unfavourable press attention and public opinion. It can be viewed as an unjust or predatory organisation, which could cause credibility and confidence to be lost.
2. Relationships with customers: Discriminatory actions might drive away and aggravate clients who feel that lending decisions are biased or unfair. Customer discontent, turnover, and harm to long-term relationships may result from this.
3. Legal and regulatory concerns: The bank may face legal and regulatory challenges because of discriminatory lending practises.

b. You can use the following to gather and curate a varied dataset to lessen bias in the AI model for loan approval:

1. Establish fairness standards: To help with data gathering, clearly specify fairness measurements and criteria, such as equal opportunity or demographic parity.
2. Gather a variety of data: Make sure a wide variety of socioeconomic and demographic parameters are represented in your dataset. To lessen underrepresentation and prejudice, this may entail gathering data from diverse people and sources.

3. Steer clear of biased variables: Take care when using variables that serve as stand-ins for protected characteristics (such as using ZIP code as a stand-in for race). In order to avoid indirect bias, eliminate or reduce these variables.
4. Audit the data: Analyze the dataset for bias by running fairness audits and bias assessments. Identify disparities in the data and their potential impact on model predictions.
5. Data augmentation: To artificially boost the representation of underrepresented groups in the data, consider data augmentation techniques.
6. Data preprocessing: To balance the dataset and lessen bias, use methods like re-weighting or re-sampling. Under sampling overrepresented groups or oversampling underrepresented groups may be involved in this.

c. By allowing the model to learn from a more representative and equitable data source, retraining the AI model with a more balanced dataset can aid in reducing prejudice and discrimination in loan approvals. This procedure may entail:

1. Relearning: Use a balanced dataset to train the model, assigning each group equal weight and reducing differential impact through the use of fairness-aware machine learning algorithms.
2. Fairness evaluation: To make sure the model no longer behaves discriminatorily, assess its performance on a regular basis using fairness criteria.
3. Difficulties: Retraining can provide a number of difficulties, such as greater data curation and gathering requirements, model retraining expenses, possible performance trade-offs, and the requirement for continual monitoring and modification to preserve equity as the data changes.

Two additional strategies the bank could implement to minimize bias in the loan approval AI system are:

1. Explainable AI (XAI): Implement XAI techniques to make the AI model's decision-making process more transparent. This can help identify and rectify bias by providing insights into which features are driving decisions and why certain outcomes occur.

2. Diverse model development: Create multiple AI models using different algorithms and approaches. By diversifying the models, you can compare their outputs and choose the fairest one, reducing the risk of bias inherent in a single model.

In order to verify that the AI system complies with fairness and anti-discrimination laws, the bank should also have a strong mechanism in place for monitoring and auditing the system's performance. Finally, the bank should develop explicit rules and procedures for fair lending and participate in regular employee training to combat bias.

Scenario 5: Transparent AI Fraud Detection Model

Scenario:
A bank has developed a sophisticated AI model for fraud detection, which has shown excellent performance in catching fraudulent transactions. However, the model's complexity makes it challenging to explain its decisions to customers and regulatory authorities.

Questions:

a. What are the potential risks and consequences for a bank if it doesn't prioritize transparency and explainability in its AI fraud detection model? How could these risks affect the bank's customer relationships and compliance with regulations?
b. Imagine a situation where a customer questions a flagged fraudulent transaction identified by the AI model. How can the bank effectively clarify the AI's decision-making process to the customer, using one of the transparency-enhancing techniques mentioned earlier?
c. How does the concept of "feature importance" contribute to improving the transparency of an AI fraud detection model? How can the bank use this concept to provide a straightforward and clear explanation of the elements that influenced the classification of a transaction as fraudulent?

Answer:

a. A bank that neglects to give openness and understandability a priority in its AI fraud detection model runs the risk of the following outcomes:

1. Mistrust from customers: If customers don't know how the bank's fraud detection system generates choices, they can grow doubtful or stop trusting it. Customers may get dissatisfied and decide to do business elsewhere as a result of this.
2. Concerns about regulatory compliance: In situations where fraud is suspected, regulatory bodies frequently demand that financial institutions give justifications for the automated judgements they make. Regulation fines and non-compliance might result from a lack of openness.
3. Legal challenges: If consumers disagree with the bank's decision-making process in the event of false positives or negatives, they may file a lawsuit against the bank, which could result in financial difficulties and legal problems.
4. Reputational harm: A bank's reputation may suffer because of unfavourable press coverage and an opaque AI model. This may influence investor confidence, shareholder wealth, and consumer acquisition and retention.

b. Researchers are trying to reverse-engineer the drivers of bias, and several approaches have been put forth to increase "explainability" and discover bias.

This entails utilising an algorithm's known inputs and outputs to determine how the algorithm decided on its own. This procedure is hard and takes time and confusing to those who are not experts approaches for machine learning, (Panch, Mattie & Atun (2019). To effectively clarify the AI's decision-making process to a customer, the bank can use techniques to enhance transparency, such as "Explainable AI (XAI)." For example, using Local Interpretable Model-

agnostic Explanations (LIME) or SHAP (SHapley Additive exPlanations), the bank can do the following:

1. Produce feature-based explanations: The bank can give the client an understandable summary of the key characteristics that affected the AI model's conclusion. The bank could clarify that the transaction was reported as fraudulent due to anomalous spending habits, inconsistency in the merchant's location, or involvement with a high-risk entity.
2. Visualise the explanation: To convey the feature importance and decision-making logic in an approachable manner, employ visualisation approaches. The relative relevance of each factor in the decision might be displayed via graphs, charts, or straightforward textual explanations.
3. Interactive interfaces: Create user-interactive interfaces that let users engage with the explanation provided by the model. Consumers can investigate the rationale behind the choice by manipulating various parameters and seeing how those changes impact the result.

By employing these strategies, the bank can increase transparency and consumer trust by giving clients a more understandable and concise explanation of why a transaction was reported as fraudulent.

c. By indicating which features or factors had the most impact on a certain classification decision, the idea of "feature importance" helps to increase the transparency of an AI fraud detection model. Using this idea, the bank can give a direct and understandable explanation of the factors that affected a transaction's designation as fraudulent by:

1. Feature ranking: The bank can indicate which factors had the biggest influence on the decision by ranking the features included in the model from most to least significant.
2. Quantifying importance: To show the relative relevance of each component in the choice, give each feature a numerical value or percentage.
3. Making the most of visualisation: To show the relevance of each characteristic, make visual representations like bar charts or feature importance plots. Customers will find it simpler to understand the main deciding elements as a result.
4. Giving explanations in plain language: Provide brief, non-technical explanations of each feature's significance alongside the feature importance rankings. For instance, "Unusual location" or "High transaction amount" might be clarified in simple terms.

This method of utilising feature importance allows the bank to offer a transparent and intuitive explanation to customers, making it clear which factors led to the AI model flagging a transaction as fraudulent.

References

OECD. (2021). *Artificial Intelligence, Machine Learning and Big Data in Finance Opportunities, Challenges and Implications for Policy Makers*. OECD. https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf

Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. *Journal of Global Health*, *9*(2). https://doi.org/10.7189/jogh.09.020318