WHITEPAPER

Explainable AI:

Building trust in business decision-making

fractal

Introduction

The ethical implications of artificial intelligence (AI) have come to the forefront in recent years. While AI has the potential to revolutionize industries and enhance process efficiency, on the one hand, there are concerns about perpetuating bias and discrimination, infringing on privacy, and making difficult-to-understand decisions on the other. The responsible AI movement seeks to develop ethical and transparent AI systems that align with societal values and respect to human rights.

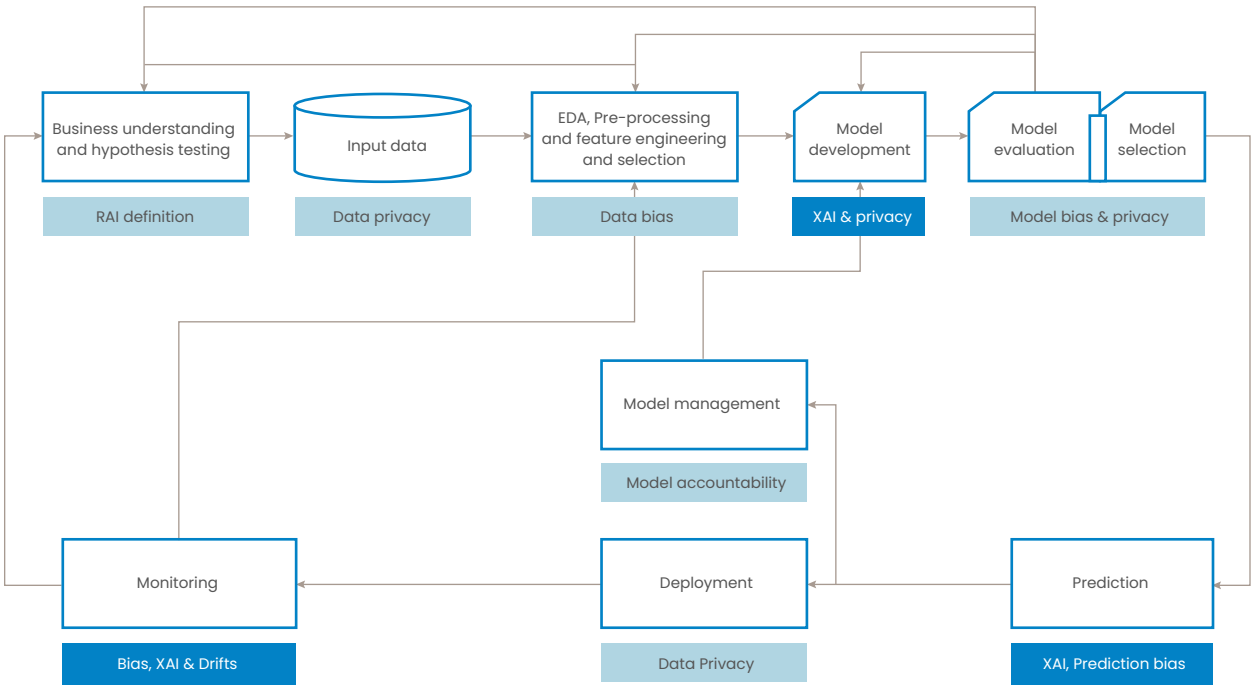
Explainable AI (XAI) has emerged as an essential component of responsible AI, potentially elevating user experience by bolstering confidence in the AI's ability to arrive at optimal decisions. It enables AI systems to provide clear and **understandable explanations of their decision-making processes.** It empowers stakeholders to understand and trust the reasoning behind the decisions and identify any biases or errors in the system's logic. Furthermore, businesses can leverage XAI to effectively manage, interpret, and trust AI systems while mitigating associated risks with deploying AI technologies.

Why Explainable AI in business decision-making?

Business leaders use AI to gain insights into complex data sets, identify trends and patterns, and make predictions that can inform strategy and drive growth. However, these decisions must be explainable and transparent to be trusted by stakeholders, including employees, customers, and investors. Therefore, incorporating explainable AI into business decision-making processes is ethical and necessary for building trust and achieving sustainable success. By providing clear explanations, XAI promotes transparency, fairness, and accountability in AI systems, ultimately leading to better decision-making outcomes for businesses with a competitive advantage.

According to a report by Research and Markets, the global Explainable AI (XAI) market was valued at USD 3.50 billion in 2020 and is projected to reach USD 21.03 billion by 2030, growing at a CAGR of 18.95% from 2021-2030.

Integrating RAI and XAI into the Data Science Lifecycle



Different Components of XAI Framework

A comprehensive XAI ensures that AI systems are transparent, auditable, and fair and helps businesses make better-informed decisions based on AI-generated insights. With this framework, enterprises can confidently understand how AI systems make decisions, the factors behind those decisions, and how to mitigate any associated risks.

Can we explain our data and its features?

- Feature sensitivity check within exploration
- Interpretable feature engineering
- Feature dependency check on target
- Feature weights calculation on target

Can we explain how agnostic model works?

- Global and local explanations
- Split and compare quantiles
- Deep and tree SHAP

Can we explain how specific model works?

- Gradient-based attribution methods
- Explanation by simplification
- GAM plots

Can we explain the risk associated to business?

- Risk monitoring assessment
- Risk calculation in probabilities/deciles
- Trade-off between accuracy and interpretability
- Counterfactuals for decision-making

Explainability for business stakeholders

In the business world, interpretability and explainability are crucial for stakeholders to understand machine learning model results and errors. This understanding helps product owners make informed financial decisions. With clear explanations provided by AI and machine learning models, users gain confidence, and developers can justify their models' validity. Transparent modeling also ensures accountability and regulatory compliance for C-suite executives. Additionally, it reduces ambiguity and promotes trust, essential to business success.

Incorporating explainability in machine learning algorithms can help mitigate risk and build trust among stakeholders, resulting in the successful adoption and application of AI technologies.
To illustrate this point, the below technique based on specific use cases can be utilized.

Split & Compare Quantiles is a valuable technique for defining decision thresholds in classification and regression problems. By enabling model evaluation and decision-making, this approach provides a clear understanding of how the model's predictions impact the business objectives, making it useful in the data science toolkit.

Salient Features of SCQ Plot

- Model Agnostic
- Business Risk Analysis
- Quantile-Based Analysis
- Data-Agnostic
- Effective Decision-Making
- Granular Risk Assessment
- Post-Deployment Analysis
- Customizable Binning

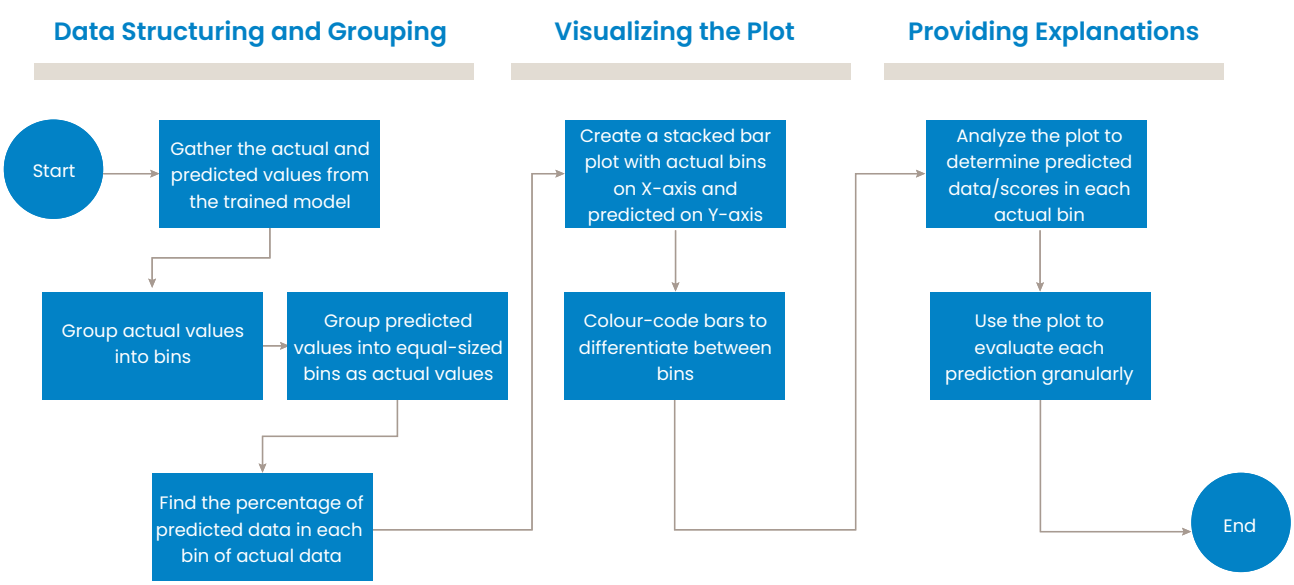
Visualizing the invisible: Mitigating business risk with split & compare quantiles (SCQ)

The Split & Compare Quantile Plot is a visualization tool that can aid businesses in decision-making and risk mitigation. This technique enables a comprehensive evaluation of machine learning models across various subsets of data, helping companies optimize their strategies and achieve their goals.

Businesses can take corrective actions and improve their overall performance by identifying areas where their models may be underperforming.

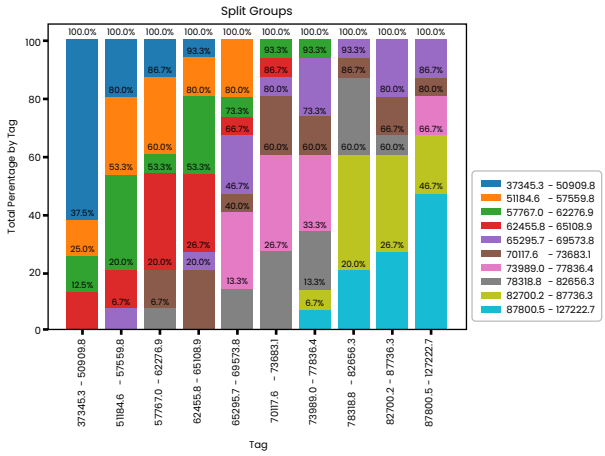
To create a split and compare quantile plot, one can first split the dataset into equal quantiles and then separate the outcomes into favorable and unfavorable categories. For instance, the sample data can be divided into deciles and categorized based on their labels. After dividing the data into equal-sized bins, the percentage of observations in each bin can be computed for both favorable and unfavorable outcomes. This approach is straightforward but efficient, enabling the analysis of a model's performance and facilitating informed business decisions.

Flowchart for Assessing Machine Learning Models with a Split and Compare Quantile Plot



Statistical Applications of Split and Compare Quantiles

SCQ for Regression Problem



Salient Features in Action

- Model Agnostic
- Quantile-based Analysis
- Data-Agnostic
- Customizable Binning
- Post-Deployment Analysis

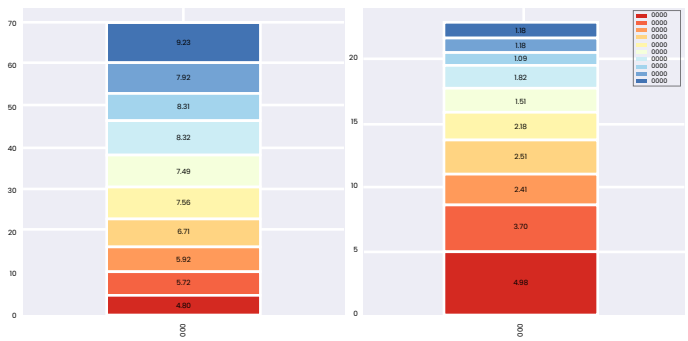
The SCQ plot can provide valuable insights into the accuracy of machine learning models post-deployment when actual values are not available.

For instance, consider a prediction of 52000 for which no actual value is available. In this scenario, the Split & Compare Quantile Plot can be used to evaluate the accuracy of the prediction and its associated risks. By examining the plot, we can see that the prediction falls into the bin marked with an orange color quantile in the legend and the 2nd quantile as per the x-axis. Further analysis reveals that in the second quantile, the orange color is accurate only ~30% of the time, while ~13% of the time the prediction falls in value bins that are lower than the actual value, and the remaining ~55% of the time, it falls in the bins above the actual value.

SCQ for classification problem

While machine learning models can be highly accurate, there's inevitably some error associated with them, potentially leading to incorrect predictions. Consider the example of identifying mortgage defaulters in financial services. Even the most sophisticated models can't achieve 100% accuracy and may misidentify non-defaulters as defaulters.

In such scenarios, stakeholders can benefit from using Split & Compare Quantile (SCQ) charts, which provide a clear picture of the degree of error associated with a model's predictions. By breaking down the data into deciles, this chart highlights all the labels within each decile, enabling stakeholders to establish the error level the model will likely have at a given threshold.



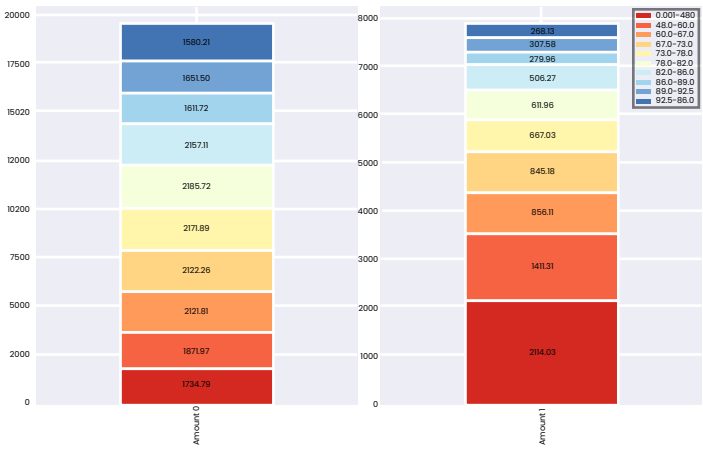
Assuming a decision threshold of 67% has been identified, which will enable the bank to service 65.25% of customers. Interestingly, this threshold will also result in 11.79% of customers being identified as having a probability of default. In such cases, stakeholders would want to minimize the risk reflected by the model by selecting a threshold that minimizes the error.

The first part represents the potential commercial loss resulting from incorrect labeling of defaulters as not likely (false positives), and the second part represents a potential loss of opportunity resulting from false negatives.

The stakeholders should use this information to determine the optimal decision boundary that minimizes false positives and negatives. By doing so, the bank can minimize its overall risk exposure while maximizing its potential revenue opportunities.

Business application of split and compare quantiles

Financial Decision-Making with SCQ



Salient Features in Action

- Post-Deployment Analysis
- Business Risk Analysis
- Effective Decision-Making
- Granular Risk Assessment

Similarly, in the SCQ plot, we can replace probabilities with corresponding monetary values. Doing so allows us to analyze the total monetary value that we will get at a chosen probability threshold. For instance, at a 67% decision threshold, the total monetary transaction value can be estimated at $17288.51K. However, there is a possibility of losing $3508.1K due to model error.

In such a scenario, stakeholders would like to select a decision threshold or bins of decision thresholds that minimize the dollar value loss resulting from model errors. By doing so, businesses can optimize their decision-making processes and minimize their overall monetary losses.

Conclusion

Traditional practices in explainable artificial intelligence (XAI) have typically been limited to basic model explainability and data visualization. However, it's vital to take things one step further, dissect the model and understand the potential risks associated with errors in the model. Surprisingly, a lower error rate doesn't necessarily equal lower financial loss and vice versa. As such, it's crucial not just to analyze the model's errors statistically but also from a monetary standpoint before making a decision on the threshold.

It is important to note that stakeholders may prefer choosing bins of different thresholds instead of a single flat decision boundary. This approach can provide a better trade-off between the monetary value of the model's errors and its accuracy.

Authors

Sakshi Gulyani
Consultant,
Fractal Dimension

Sray Agarwal
Principal Consultant,
Fractal Dimension

Supriya Panigrahi
Consultant,
Fractal Dimension

About Fractal

Fractal is one of the most prominent providers of Artificial Intelligence to Fortune 500® companies. Fractal's vision is to power every human decision in the enterprise, and bring AI, engineering, and design to help the world's most admired companies.

Fractal's businesses include Crux Intelligence (AI driven business intelligence), Eugenie.ai (AI for sustainability), Asper.ai (AI for revenue growth management) and Senseforth.ai (conversational AI for sales and customer service). Fractal incubated Qure.ai, a leading player in healthcare AI for detecting Tuberculosis and Lung cancer.

Fractal currently has 4000+ employees across 16 global locations, including the United States, UK, Ukraine, India, Singapore, and Australia. Fractal has been recognized as 'Great Workplace' and 'India's Best Workplaces for Women' in the top 100 (large) category by The Great Place to Work® Institute; featured as a leader in Customer Analytics Service Providers Wave™ 2021, Computer Vision Consultancies Wave™ 2020 & Specialized Insights Service Providers Wave™ 2020 by Forrester Research Inc., a leader in Analytics & AI Services Specialists Peak Matrix 2022 by Everest Group and recognized as an 'Honorable Vendor' in 2022 Magic Quadrant™ for data & analytics by Gartner Inc.

For more information, visit fractal.ai

Corporate Headquarters
Suite 76J,
One World Trade Center, New York,
NY 10007

Get in touch