Security and Governance for AI Models

Introduction

In today's rapidly evolving digital landscape, artificial intelligence (AI) and machine learning (ML) models have become integral to the operations of many organizations. However, with great technological advancements come great risks. It's imperative for businesses to address the vulnerabilities and challenges associated with AI, and that's where our Model Risk Management system steps in as your frontline defense against adversarial AI threats.

In addition to identifying and mitigating risks, it's crucial to evaluate models on key criteria to ensure their reliability and ethical use. Our Model Risk Management Service places a strong emphasis on evaluating models in terms of robustness, fairness, and explainability.

We developed a solution that has a battery of attacks to test the ML model for its robustness. The same tool also enables data scientists to test the explainability and fairness of the model.

Implementation

To ensure accessibility and scalability, our infrastructure comprised the following tools:

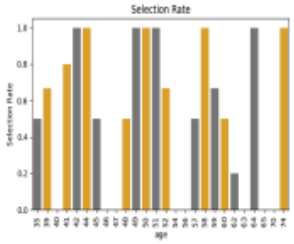
FastAPI Endpoints: We developed a set of APIs through FastAPI endpoints, allowing data scientists to expose their adversarial attacks simulations, fairness evaluations, and model explainability analyses. This facilitated easy access and seamless integration into other systems.

AWS S3 Bucket: Models and datasets, stored in the efficient .joblib format, were securely maintained in an AWS S3 bucket. This ensured streamlined data storage and retrieval for machine learning experiments.

Development Environment: PyCharm served as the integrated development environment (IDE) for coding FastAPI endpoints. Dockerdash was utilized for containerizing the code, enhancing portability and ease of management.

Graphs Generated

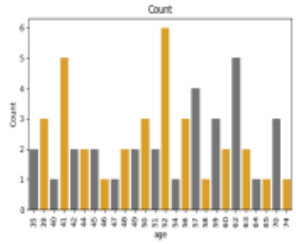
Age selection rate comparision



2023-09-07 19:55:49.332775

full size

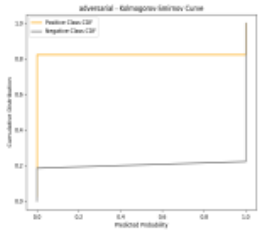
Age count comparision



2023-09-07 19:55:49.332775

full size

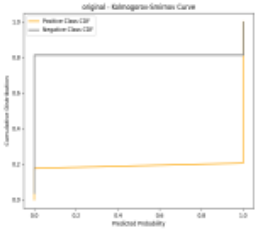
Ks curve adversarial



2023-09-07 19:55:49.332775

full size

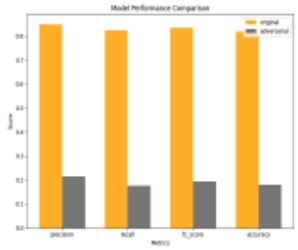
Ks curve original



2023-09-07 19:55:49.332775

full size

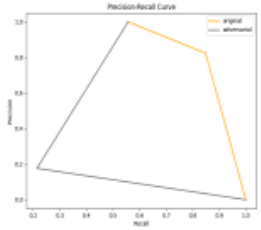
Precision graph



2023-09-07 19:55:49.332775

full size

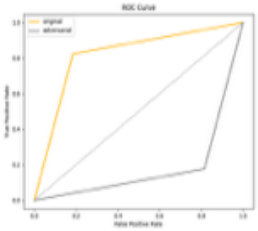
Precision recall curve



2023-09-07 19:55:49.332775

full size

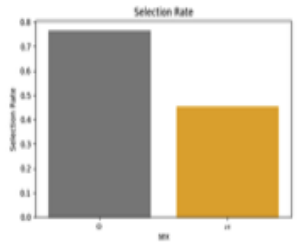
Roc curve



2023-09-07 19:55:49.332775

full size

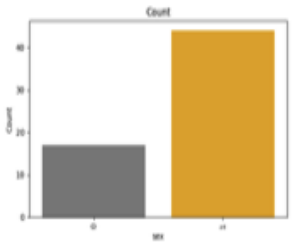
Sex selection rate comparision



2023-09-07 19:55:49.332775

full size

Sex count comparision



2023-09-07 19:55:49.332775

full size

Impact

The solution served as a valuable tool to validate model risks, ensuring that the AI ecosystem remains secure and reliable.

In an era where AI plays a pivotal role in business operations, the security and reliability of AI models cannot be compromised. The Model Risk Management service offers a robust and holistic approach to AI security, enabling organizations to harness the power of AI while safeguarding against potential threats.

As the AI landscape continues to evolve, our Model Risk Management system stands as a reliable partner in ensuring that enterprise AI/ML models remain secure, compliant, and resilient in the face of adversarial challenges.