LATEST VERSION:

**Intro: - Noor**

Hello and good afternoon to everyone and welcome to our presentation. We are the EY Data & AI Team, and we worked on a Generative AI solution for NRMA Insurance to assist claim agents in identifying + processing fraudulent insurance claims. We’re excited today to present our innovative and comprehensive solution.

We will start off with a welcome to the country, if Josh/Caio could please assist with that.

Thankyou Josh/Caio for acknowledging the traditional owners of the land.

Now, to start with a brief team introduction of the data team, I’m Noorullah Khan, the Project Manager, Lead Data & Software Engineer. Alan is the Cloud Solutions Architect, Adam is the Data Engineer and Modeller, and Ninuri and Tashiya are our Business Analysts. From the AI team we have Dan who is our software engineer and Eugene the AI and Machine Learning Engineer. There were some members of the AI team who were unable to attend today due to conflicting schedules.

**Synthesising - Noor**

We used 2 data synthesising methods to generate a total of 8000 rows which is about 38% of the total dataset. These rows were split evenly between 2 data generation methods, each with their own unique approach.

Method 1 generated 4000 rows using remapping and mathematical sampling. This works by remapping the existing data and then applying mathematical sampling techniques to generate new rows which fit to the structure of the original data. It preserves the authenticity of the fraud column by ensuring it remains untouched by the remapping process. Method 1 also allows for quick and efficient data generation and is ideal when needing to generate data quickly with minimal resources. It has a low computational cost.

Now onto Method 2 which also generated 4000 rows but using CTGAN, a more advanced approach. CTGAN is a sophisticated machine learning method based on neural networks. It learns the underlying patterns and complex relationships in the data. A key strength of this method is its ability to capture intricate relationships between variables – particularly the fraud column – making it an ideal method for accurate and high-quality data generation. CTGAN is best suited for handling complex data structures where it is critical to ensure realism i.e. generated data which can replicate the real-world patterns.

**Data Merging - Noor**

During the merging process, we combined three (cleaned and enriched) original datasets and two synthetic datasets, resulting in a final dataset of over 20,000 rows, or more precisely, 20,877 rows.

This includes 1,000 rows from Dataset 1, about 1,500 rows from Dataset 2, about 10,000 rows from Dataset 3, and 4,000 rows each from the two synthetic datasets.

To maintain consistency and have a form of identifier, we assigned a unique index for each row. Each row also has a fraud label which is a numerical boolean, with 1 representing fraud and 0 representing not fraud which is crucial for ML & AI training & testing.

The merged dataset is now ready for the AI team to finetune their model training.

**WebApp Integration - Noor**

So carrying on from the AI team’s model, how can NRMA claim agents actually use it? For this we developed a comprehensive front end Web App and claim processing system.  
In essence, the data is used to make a trained AI model which then via API’s connects to the WebApp. An independent table of current & live insurance claims is also linked to the WebApp.

The claim agent uses the WebApp to investigate if a claim is fraudulent with the assistance of Generative AI. Based on his findings he can make claim processing actions like approving or escalating and adding notes. His interactions with the AI and claim processing system will be saved and logged.

**Live Demo - Noor**

Now let’s have a live demo of the fraud detection system UI and web app.

Main points:

* Secure Login
* Dashboard Overview
* AI-Powered Fraud Detection
* Claim Management Features
* History & Insights page
* User Friendly Interface