World Wide Technology

Make a new world happen

Aurum Mining Corporation (AMC)
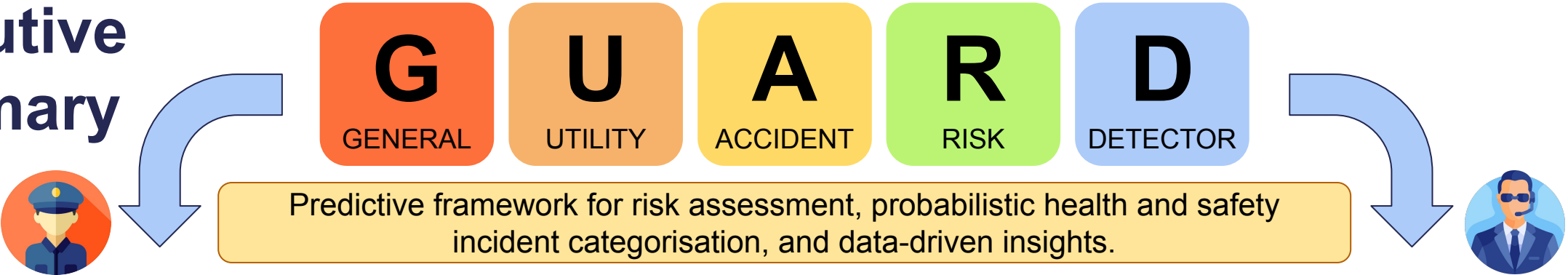
Empowering the Health & Safety organization

Trilytics'23 Conclave Case Study

Team Maverick Minds

Susanket Sarkar
Yash Jangade

Executive Summary

G GENERAL **U** UTILITY **A** ACCIDENT **R** RISK **D** DETECTOR

Predictive framework for risk assessment, probabilistic health and safety incident categorisation, and data-driven insights.

ACCIDENT PREDICTOR

- Data Operator present at mine inputs data into model present in their local machine.
- Model processes and predicts grey areas

- Health and Safety team takes precautionary measures (based on model prediction)
- Reduces risk of accident occurring

BUSINESS IMPACT PREDICTOR

- In case an accident happens team present at site inputs data into a server.
- Model processes the data and predicts it's impact.

- Findings are delivered to AURUM's Operations Team.
- Team takes necessary measures based on model predictions to mitigate losses.

BUSINESS VALUE

SAFETY
Proactive Risk Mitigation at early stage, Cultivate Safety culture at workplace
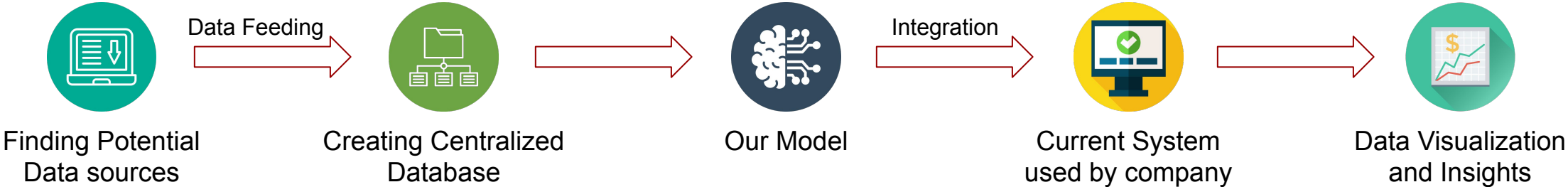
COST
Resource Optimization, Reduction in expenses related to medical, compensation, and downtime costs

DECISION MAKING
Facilitates Data Driven Decisions, quicker and accurate Incident response

PRODUCTIVITY
Minimizes accidents and days lost, leading to higher workforce productivity

EXECUTION

Finding Potential Data sources → *Data Feeding* → Creating Centralized Database → Our Model → *Integration* → Current System used by company → Data Visualization and Insights

Approach Towards Problem

Raw Data *(36016 rows x 49 columns)*

Data Preprocessing

Feature Engineering I

- Decomposed `ACCIDENT_DT` into 'Accident Year', 'Month', 'Day of Week', and 'Quarter'
- Helps to capture seasonality or Time based trends that might be present in accident occurrences.

Feature Engineering II

- Created 'BEFORE_MINE_EXPER' & 'BEFORE_JOB_EXPER' from 'JOB_EXPER', 'MINE_EXPER', and 'TOT_EXPER'
- New Features captures years worked in other mines and jobs.

Textual Data

- 'Removing Stop Word' i.e. eliminating words that don't add much meaning.
- Next, transforming the words into numerical vectors using the Gensim library (Grouping similar words together)

Standardisation

Scaling and transforming data to a common range, ensuring consistent and optimal model performance across features.

Imputation

Employed KNN Imputer to fill in missing values, which predicts these values based on record similarity.

Solution   **Accident Prediction Model**   **Impact Prediction Model**
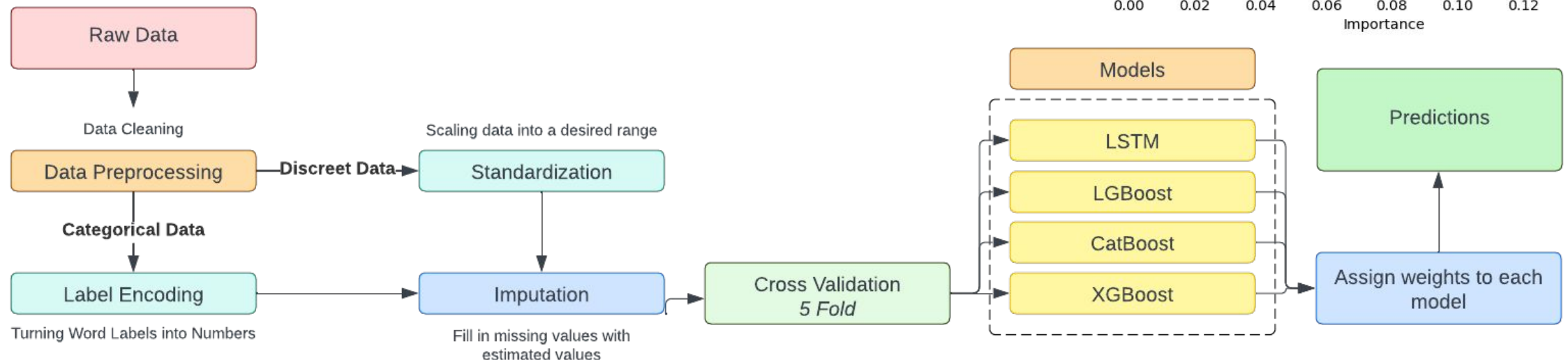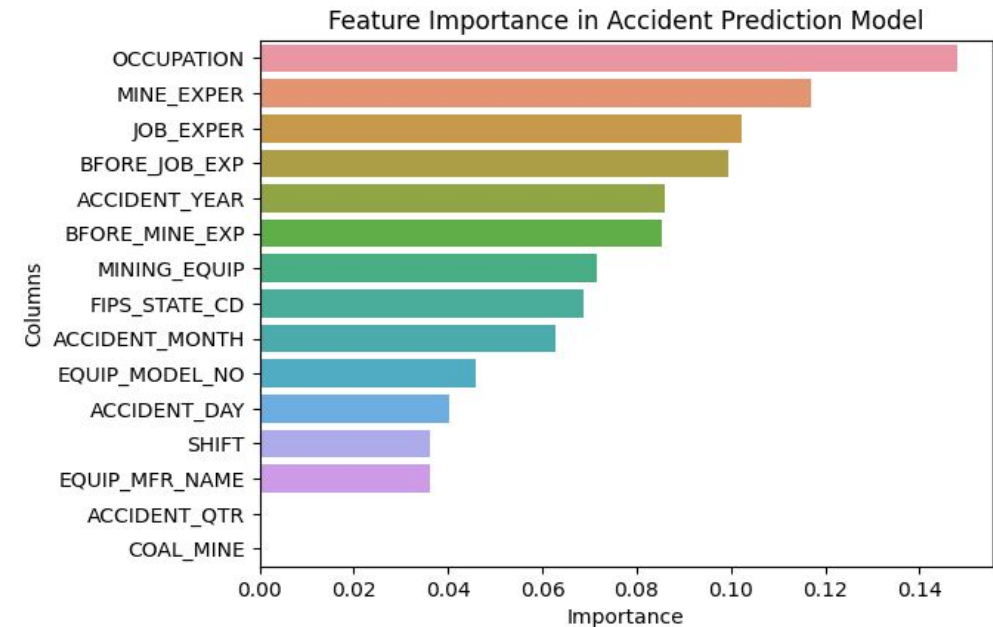
Accident Prediction Model

We created a model capable of **predicting health and safety incidents** in the mining industry. This model uses incident descriptions to **forecast whether an incident is likely to occur.**

Model Selection:

- **Sequential Data Modeling** - **LSTM**: Recognizing the **sequential nature of our data**, with events unfolding over time, we decided to utilize a Long Short-Term Memory (LSTM) model. LSTMs are a type of Recurrent Neural Network that are well-suited to learning from time-series data due to their ability to remember and utilize context from earlier points in the sequence.
- **Tree-Based Models** - **XGBoost, CatBoost, and LightGBM**: Alongside the LSTM, we also trained a selection of tree-based models: XGBoost, CatBoost, and LightGBM. These models were chosen for their **compatibility with tabular numerical data** like ours. They are known for their ability to handle mixed data types, capture complex interactions between features, and provide robust performance even with missing values or outliers.

Feature Importance in Accident Prediction Model

Impact Prediction Model

Our next model is capable of **predicting the business impact** of an mining accident. This model uses incident descriptions to **evaluate business impact and classify accident type.**

Model Selection:

- With our data prepared, we trained a diverse set of models - **LSTM, SVM, XGBoost, CatBoost, and LightGBM** - to predict the business impact of an accident. This selection combines models that are well-suited for numerical data (XGBoost, CatBoost, and LightGBM), a model good at handling text data (LSTM), and a model (SVM) known for its effectiveness on a wide range of data types**.**

- Using a diverse set of models (LSTM, SVM, XGBoost, CatBoost, and LightGBM) for accident impact prediction brings advantages: **robustness** by reducing biases, **improved accuracy** with **ensemble voting**, knowledge sharing through transfer learning, dynamic model selection, **enhanced interpretability**, and future-proofing.
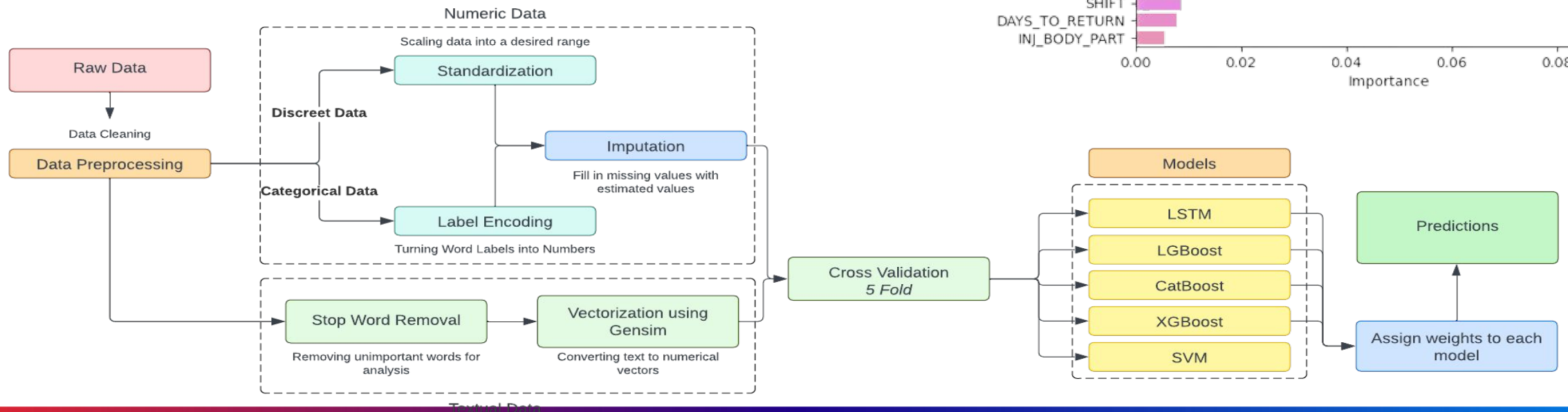
Predict, Initiate & ProTecc

Model predicts business impact, furnishes investigation triggers & help in preventing future accidents

Feature Importance in Business Impact Prediction Model
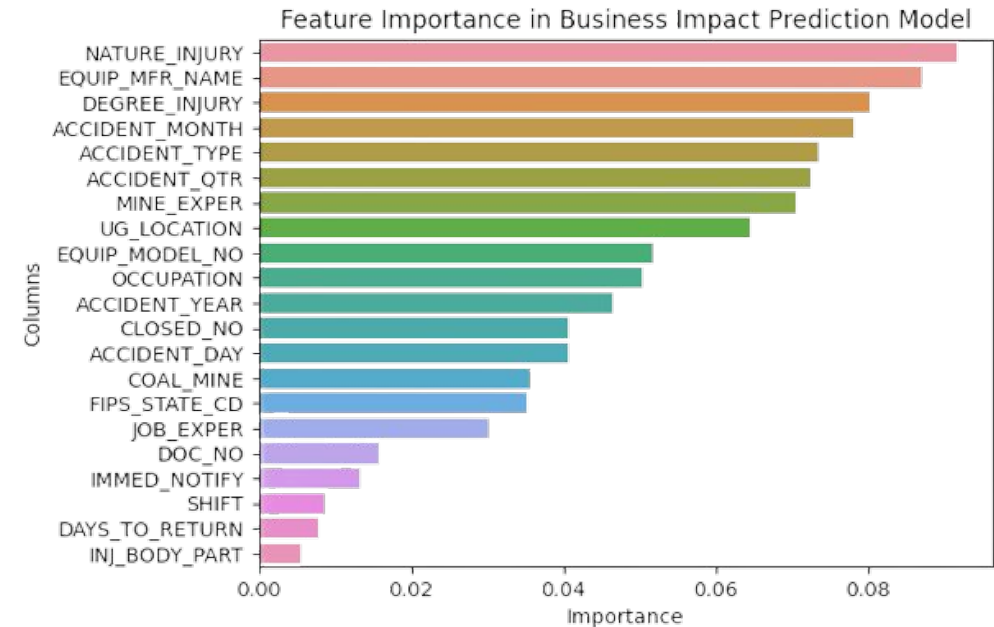
Quality Checks: Metrics & Validation

Assessed the performance of the model using cross validation technique, ensuring the model works well on different subsets of data and verifies that our model generalizes well to new data by utilizing various metrics like **Accuracy, Recall, and Mean Squared Error**
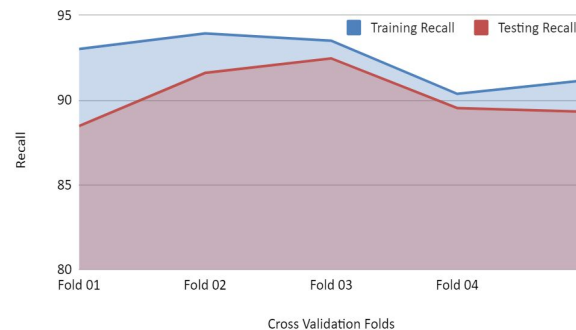
⚠️ Accident Prediction Model ⚠️

Recall is important for accident prediction models because it **minimizes false negatives**, which could lead to injuries or deaths. **In our context, a false negative means predicting no accident when an accident actually occurs.**

How the Recall changes for different validation sets

- Training and testing recall values are closely aligned. This indicates that the **model does not overfit** the training data.
- **High values for recall, precision, and accuracy** confirm that the model has been trained effectively.
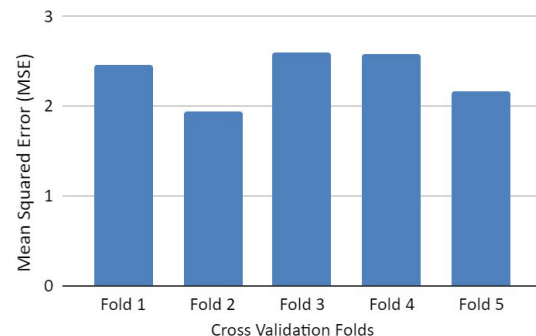
87%
ACCURACY

91%
PRECISION

92%
RECALL

- The **Mean Squared Error penalizes larger errors** between actual and predicted values more heavily.
- Our model **achieved a low MSE score**, which indicates that it was **able to make accurate predictions** on both the training and testing sets

How does the regression metrics look for the model?

💼 Impact Prediction Model 💼

- The choice of Accuracy **aims to ensure our model's ability to precisely predict** business impacts.
- **RMSE**, on the other hand, provides an estimate of how far our predictions deviate from the actual values, measured in the same units as the target variable.
- As we can see from the figure below, the LSTM model was overfitting over 40 epochs. So we only **considered upto 40 epochs.**

Accuracy of the Bussiness Impact Prediction Model
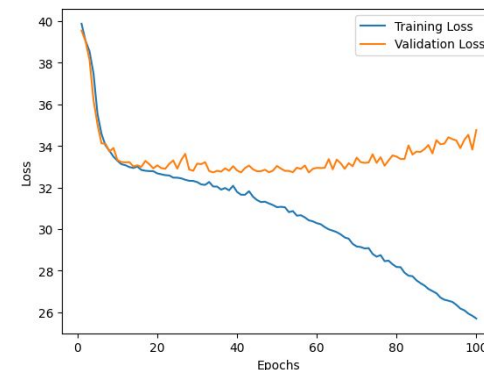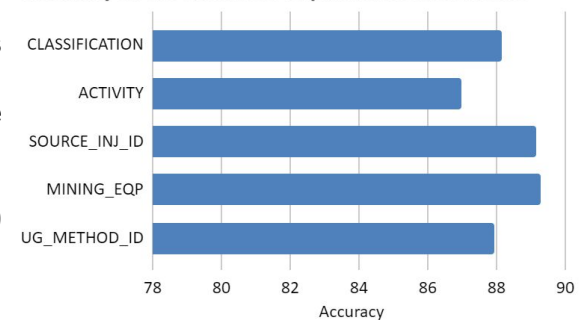
RMSE of the Business Impact Prediction Model
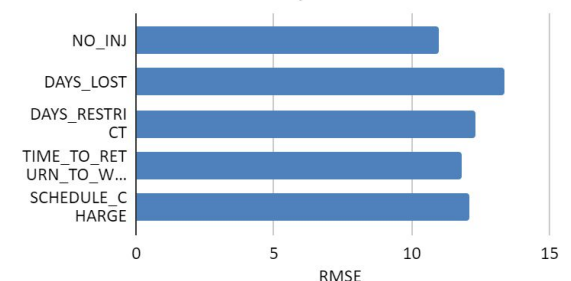
88%
ACCURACY

89%
PRECISION

87%
RECALL

11 RMSE

Our model's predictions are on average 11 units off the values

Streamlining Health & Safety Data Integration

Data Operator

- Collating information from various sources.
- Assisting in generating reports

Dashboard for Predictive Safety

- Location-specific Accident Probability
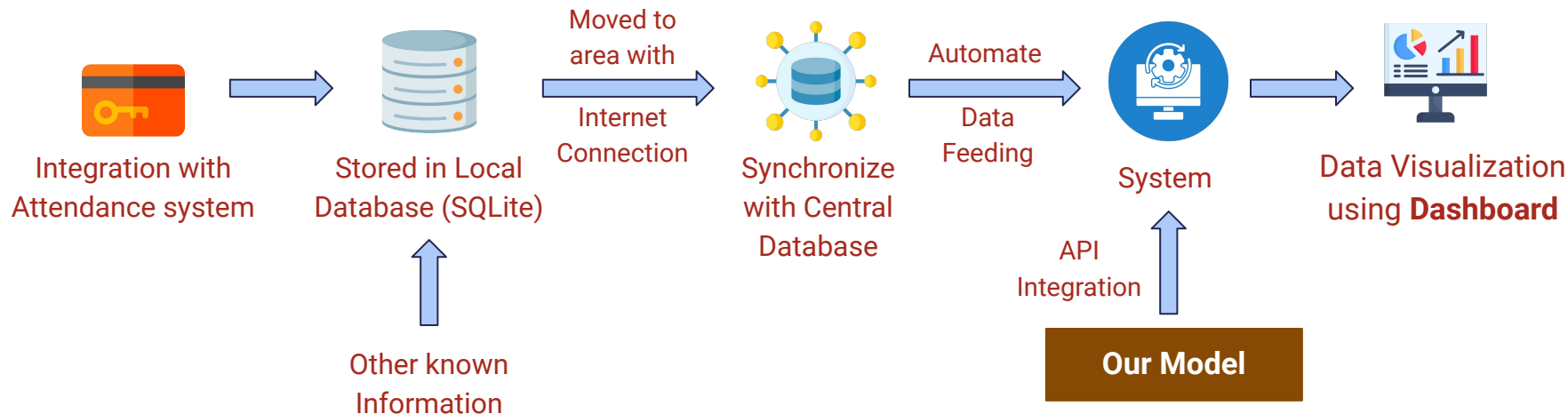- Predicts Injury Severity and Nature

Early Precautionary Measures

- Enhanced Safety Culture
- Reduce severity & frequency
- Safeguarding lives and resources

In Case Accident Happens

Accident occurs

Post-Accident Analysis Dashboard

- Accident Impact Analysis
- Incident Documentation
- Investigation Triggers
- Instant Metrics & KPIs

Easy Integration and Automation:

Integration with Attendance system → Stored in Local Database (SQLite) → Moved to area with Internet Connection → Synchronize with Central Database → Automate Data Feeding → System → Data Visualization using **Dashboard**

Other known Information

API Integration

Our Model

Tracking Incidents
Ideal Benchmarks vs. Real Occurrences

Model's Performance Metrics ⬌ Safety Benchmarks & Standards

LTIFR, TRIR

Goals & Objectives
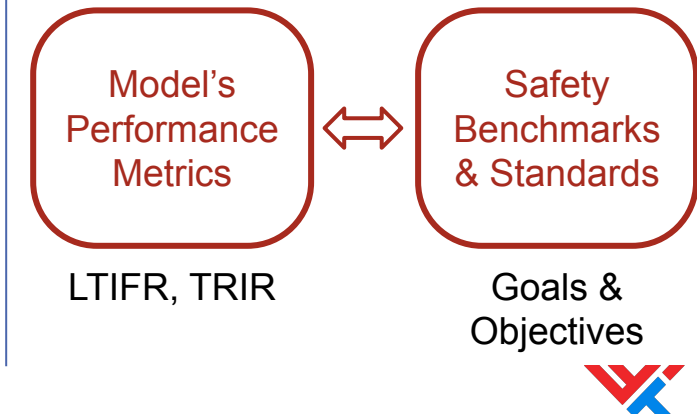
Appendix: Dashboard

A comprehensive dashboard empowering Safety and Data Management team with actionable insights for enhanced operations and risk mitigation.
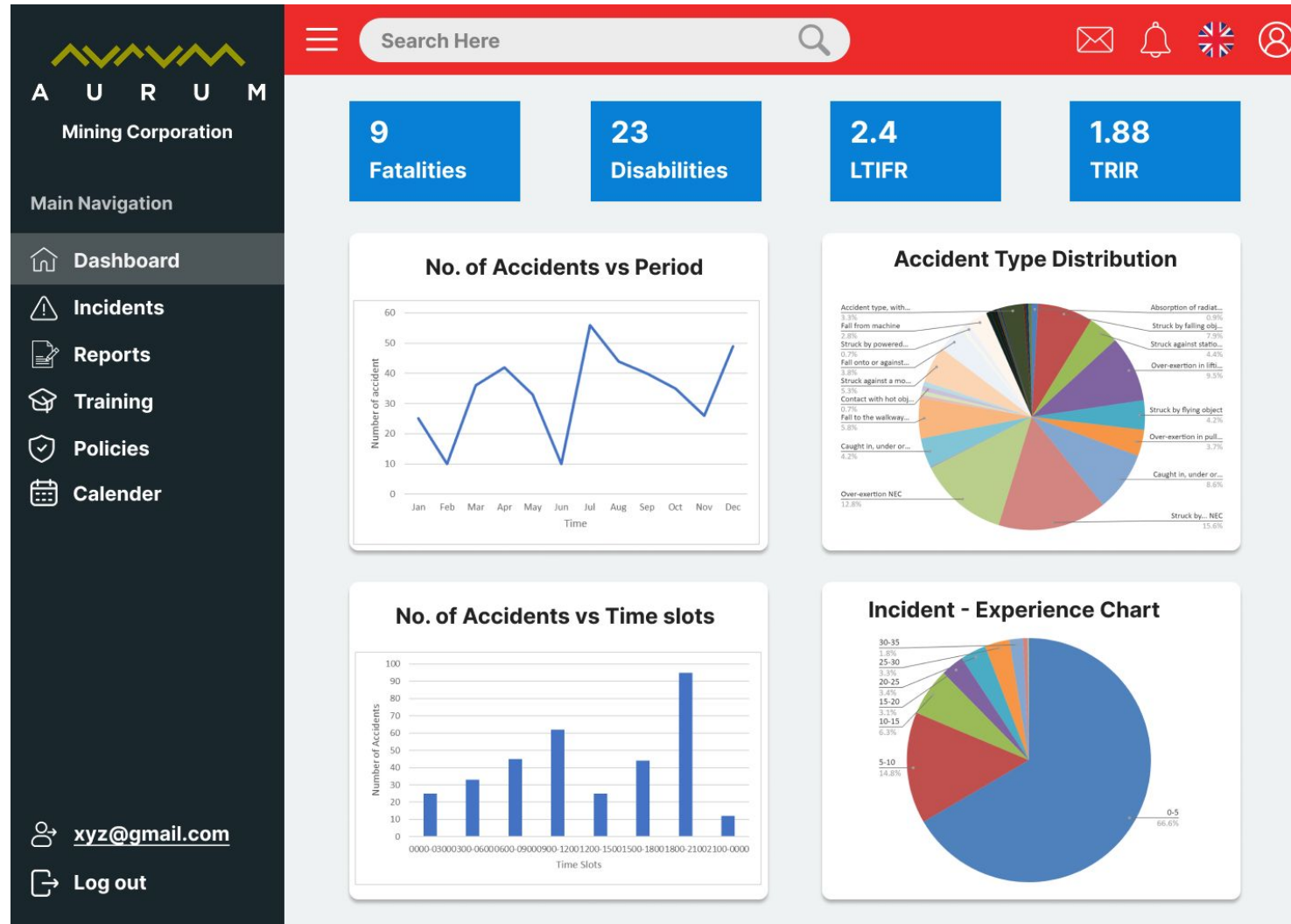
Seamless Management

- Single point access to various areas
- Easy Performance Monitoring

Centralized System

- Easy access of data to officials
- Streamlining the reporting processes

Remote Access to the Data

- Flexibility to review data on-demand
- Quick & informed decisions remotely.

Instant KPIs

- Assess alignment with goals.
- Early problem detection
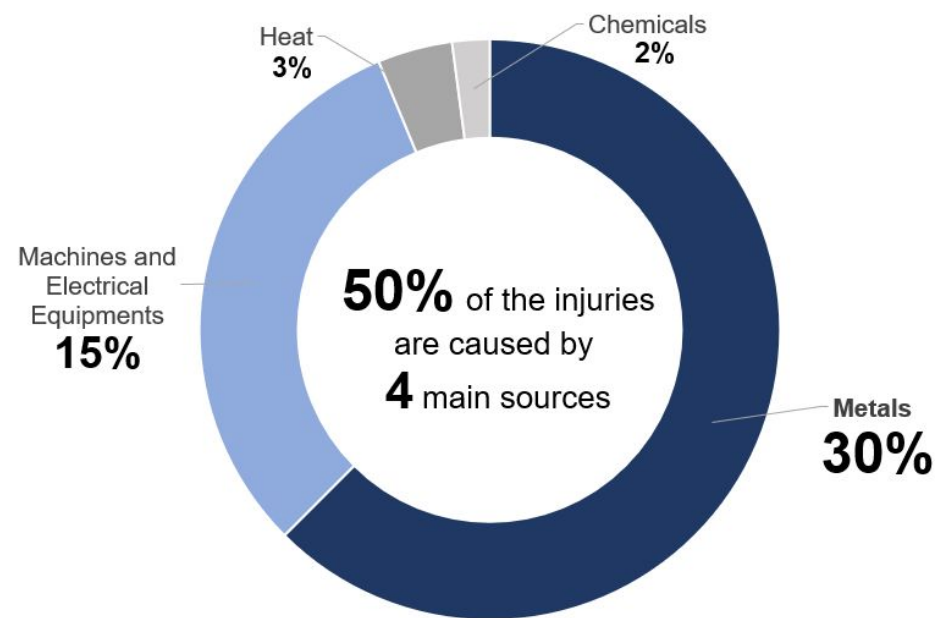
Integration with Incident Database

- Real time Insights
- Dynamic plots
- Compliant Incident records

Data Visualization

- Concise presentation of data
- Show patterns and trends
- Highlight Focus Areas

AURUM
Mining Corporation

Main Navigation

- Dashboard
- Incidents
- Reports
- Training
- Policies
- Calender

xyz@gmail.com

Log out

Search Here

| 9 Fatalities | 23 Disabilities | 2.4 LTIFR | 1.88 TRIR |

No. of Accidents vs Period

Accident Type Distribution

No. of Accidents vs Time slots

Incident - Experience Chart

Risk Analysis & Mitigation

Appendix

55% of the accidents heppens in the FILO (**F**aces, **I**ntersection and **L**ast **O**pen crosscut)

Accidents in **FILO** causes **60%** of days lost

Accidents in **FILO** causes **56%** of injuries

55% Accidents in **Underground** are caused in **FILO**

Continous Mining in **FILO** causes **40%** accidents

FILO records **45%** of all the injuries

50% of the injuries are caused by **4** main sources

- Heat 3%
- Chemicals 2%
- Machines and Electrical Equipments **15%**
- Metals **30%**

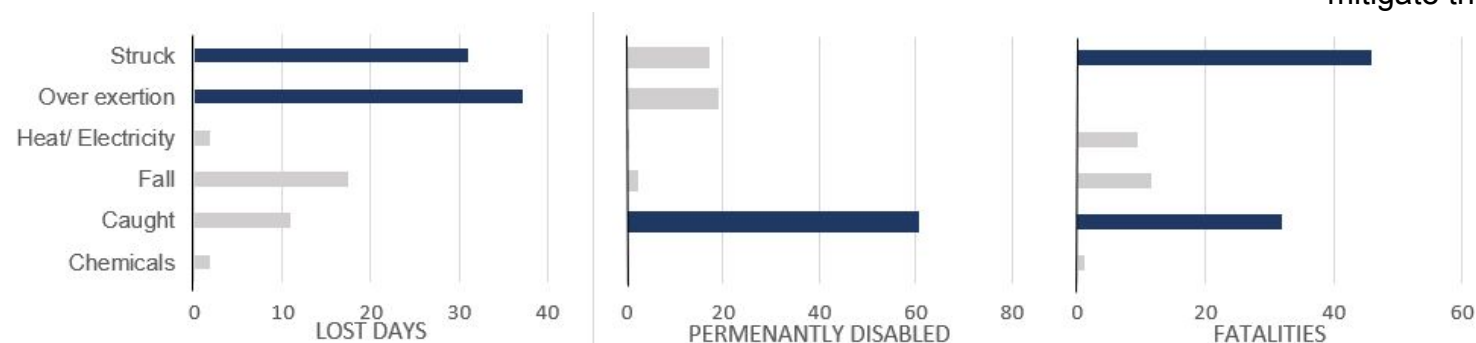**UnderGround Location:** Collectively around **55%** of the accidents happen in FILO (Faces, Intersection and last Open Crosscuts). **Regular Inspections, Proper Illumination and Safety Training** can help reduce the chances of these accidents.

Injury Source: **4 main sources** broadly contribute to 50% of the injuries. A focus on improving safety protocols, regular maintenance and inspections,and deploying protective gear can help mitigate these accidents.
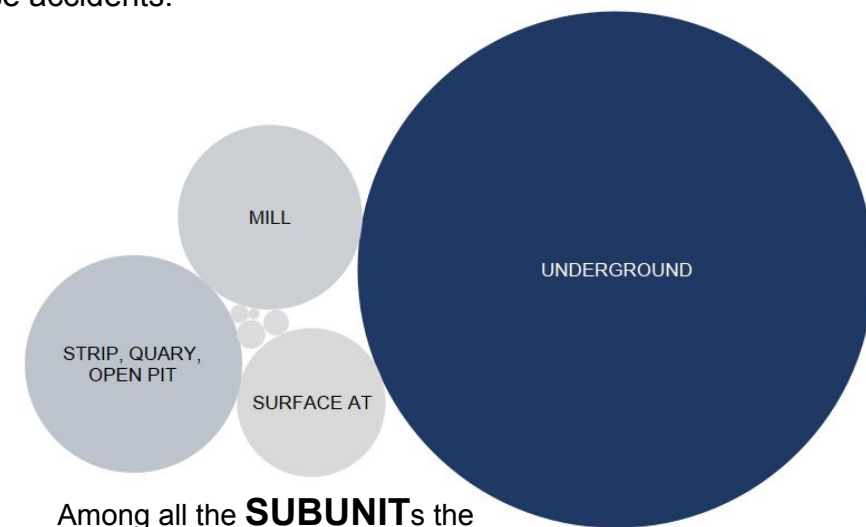
LOST DAYS (Struck, Over exertion, Heat/ Electricity, Fall, Caught, Chemicals)

PERMENANTLY DISABLED

FATALITIES

38% Lost Days caused by Over-exertion

60% of cases of Permanent Disability caused by being caught between rocks

45% of Fatalities caused by being Struck

Improved Head Gear can help reduce chances of accidents. Overexertion can be mitigated by **better scheduling** and **job rotation**

MILL

UNDERGROUND

STRIP, QUARY, OPEN PIT

SURFACE AT

Among all the **SUBUNIT**s the Underground account for **70%** of the accidents.