# Application of Supervised Machine Learning for Gold-Silver Throughput, Recovery, and Mine Production in a Philippine Gold Mine

JOSE NORBIEL G. FLORENDO

AIM - PGDAIML

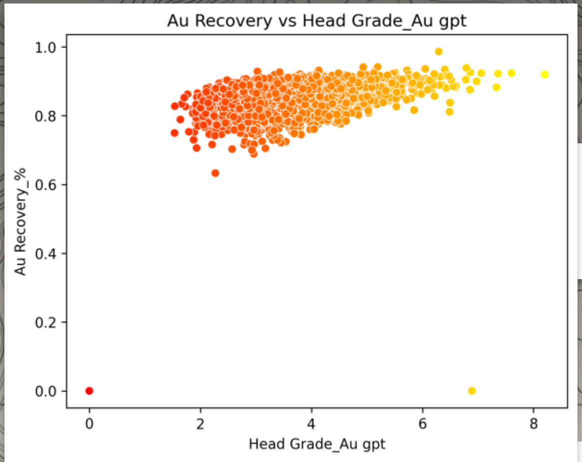
# Outline

- Business Context and Problem Statement
- Methodology
- Model Creation
- Results and Discussion
  - Comparison of Tuned and Baseline Models
- Conclusion and Recommendation

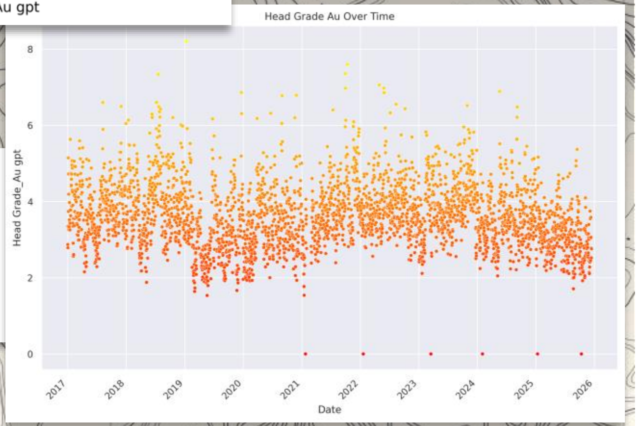
# Business Context

- A mining company's financial profitability and viability depends almost entirely on its level of output (productivity) and its output efficiency.
- Productivity and Output Efficiency in a Gold-Silver mine is measured respectively using **Gold and Silver Output** and **Recovery** (for Gold and Silver)
  - Recovery (%) is the total metal output successfully extracted from the total metal content of the ore input
    - $\frac{\text{90g gold extracted from 1 ton of Ore}}{\text{100g gold concentration in 1 ton of Ore}}$ = 90% Gold Recovery
  - Mine Tonnage is the total amount of ore (in tons) produced by a mine for a specific period.
- For a mining company to be profitable and efficient it must be able to maintain a high enough level of productivity and output efficiency given a wide range of inputs.

# Problem Statement




Gold Recovery is not linear




Gold Grade is not constant over time

- **Business problem:**
  - Recovery and Output can only be calculated **after the metal** (gold and silver) **has already been extracted** from the ore.
  - Recovery prediction is **back-calculated through Linear Regression**, which may not have sufficient predictive power for non-linear behaviors and to handle varying ore grades.
- **Business opportunity:**
  - If the mining company can **anticipate** and **adapt** swiftly by **accurately predicting key mining outputs**: [1] gold and [2] silver recovery, [3] the total ounces of gold and [4] silver produced, and the [5] using minimal inputs, the mining company can **change operational parameters immediately** to increase ore recovery, and this ultimately **maximizes value** of its raw product.

# How did we make our models?

### Data Loading & Inspection
- Detect Nulls
- Detect Datatypes

### Data Cleaning
- Drop Nulls
- Normalize data
- Outlier management

### Exploratory Data Analysis
- Visualize trends and relationships
- Feature Selection

### Scaling Data, Splitting Data for Training & Validation
- Standardize data
- Split to Train (80%) and Testing (20%) splits

### Create Baseline Models
- Linear Regression
- XGBoost Regression
- LightGBM Regression
- CatBoost Regression
- Decision Tree Regression
- Support Vector Regression
- MLPRegressor (Artificial Neural Network)

### Hyper-parameter Tuning

### Comparative Model Selection
- Plot and compare model performances

### Model Export
- Export Baseline Models
- Export Tuned Models

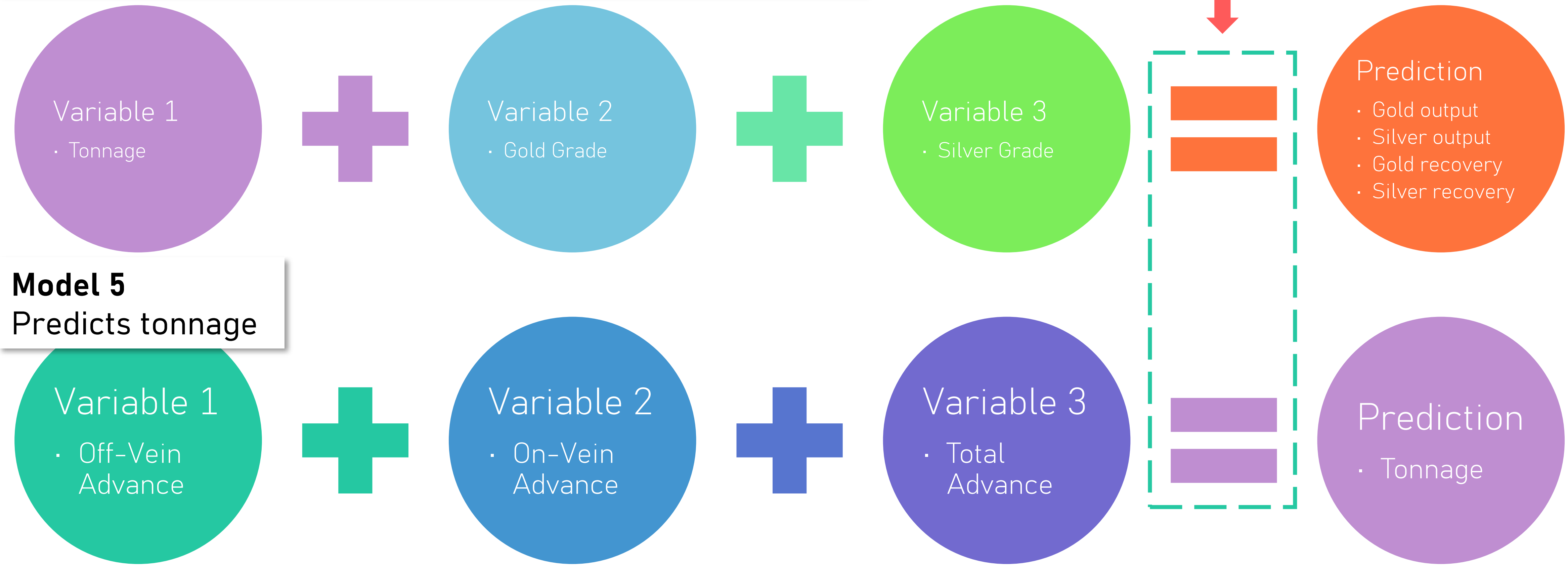
### SHAP Insights

### Discussion and Conclusion

# Where does ML/AI fit here?

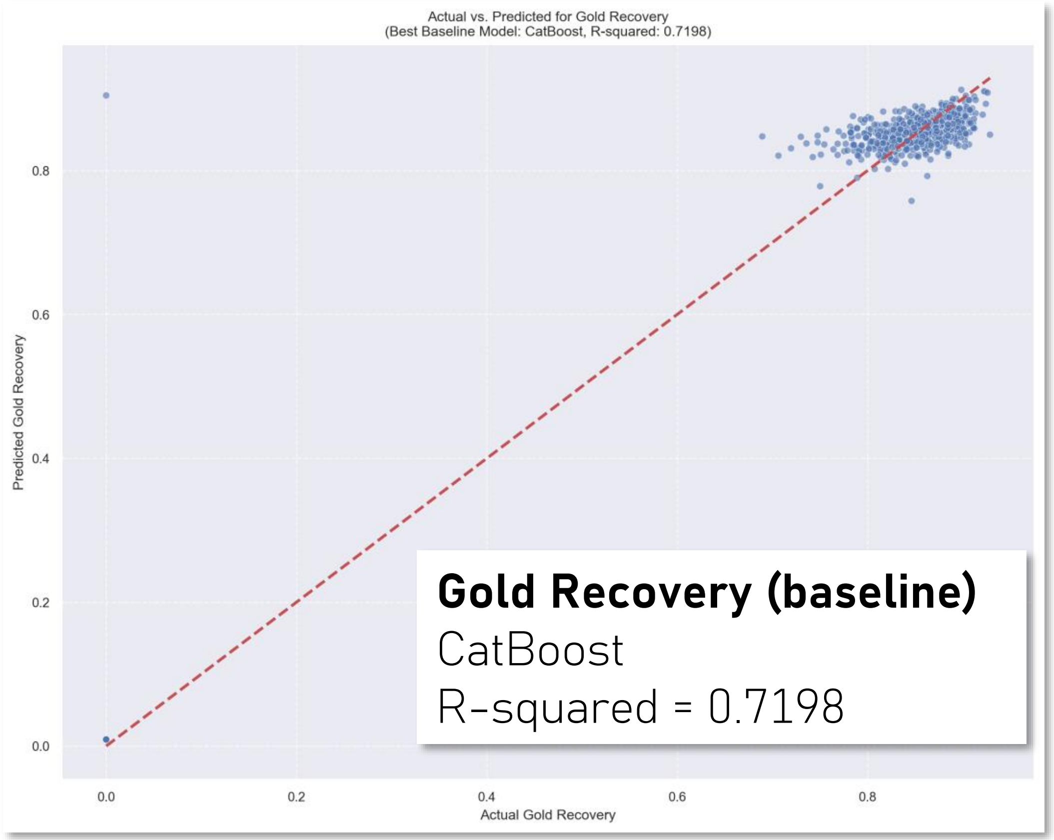
**Model 1, 2, 3, 4**
Predicts Au output and recovery, Ag output and recovery

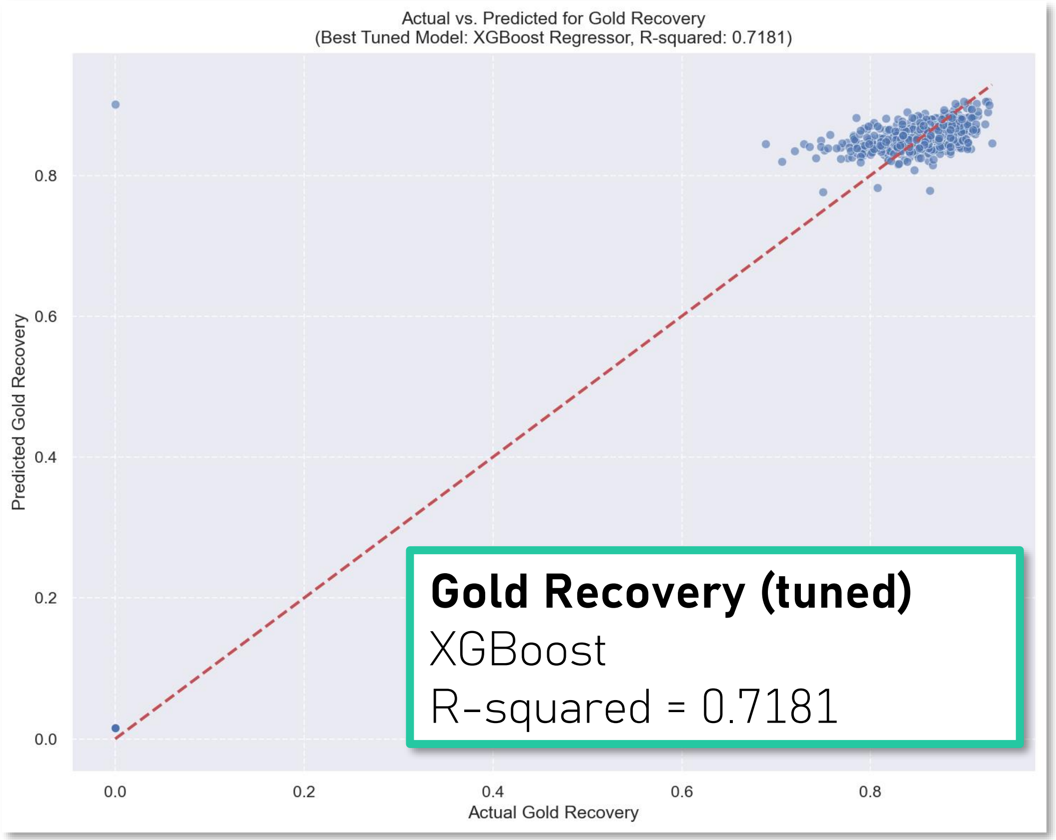
**Data Processing**
**Machine Learning**
(7 ML/AI models)




**Model 5**
Predicts tonnage

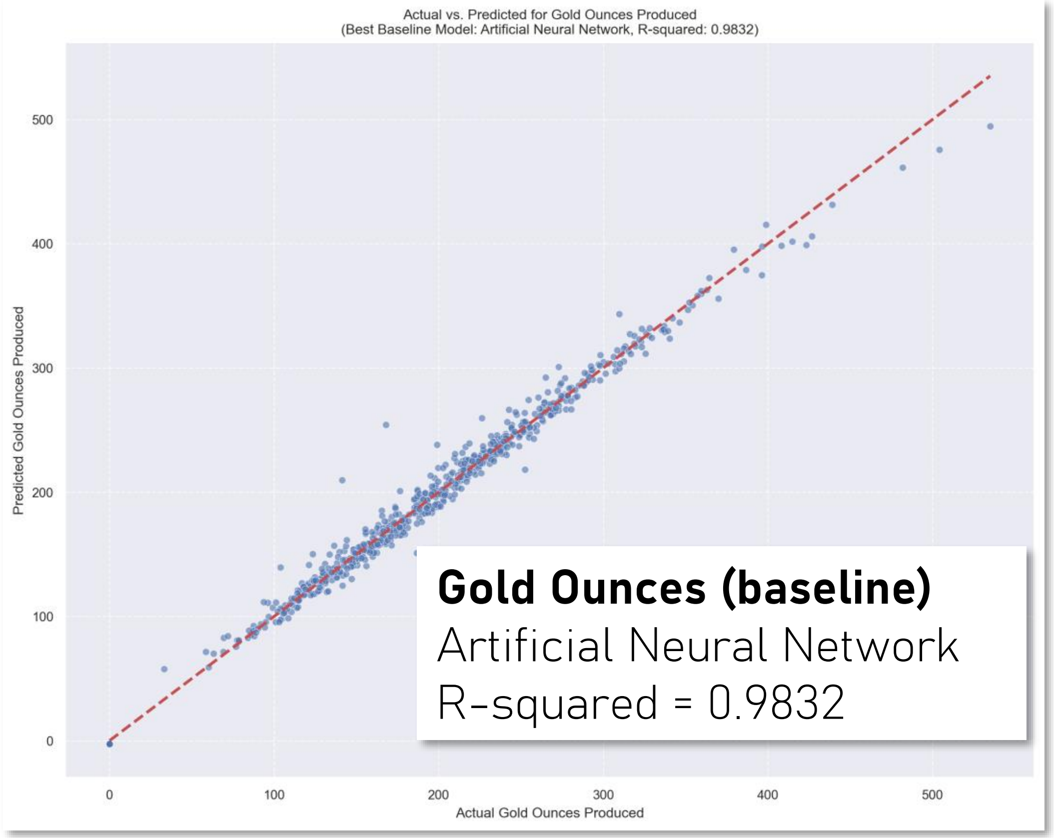
# What did we learn? – Gold Recovery



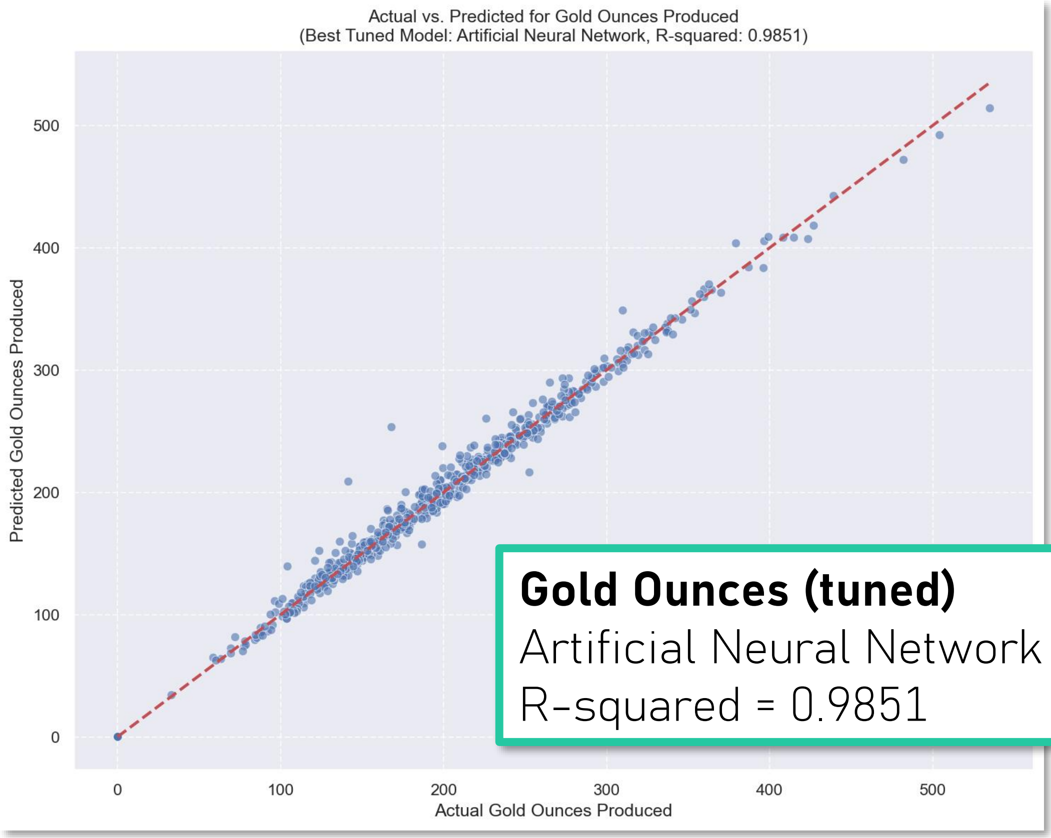




Using only Tonnage and Au, Ag grades, our model can explain 71.8% of the variation of our Gold Recovery. We are also able to predict with an error of ±4.71%. This means, we need other variables to better predict Gold Recovery.

# What did we learn? – Gold Ounces



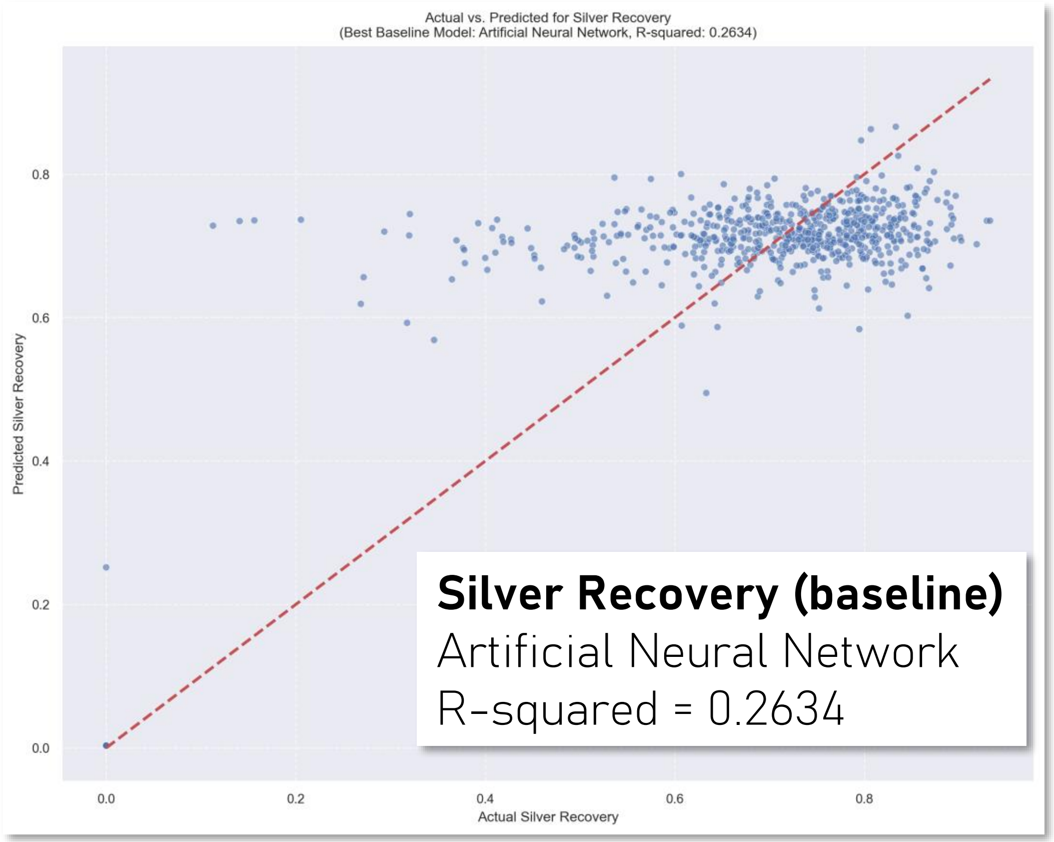

**Gold Ounces (baseline)**
Artificial Neural Network
R-squared = 0.9832



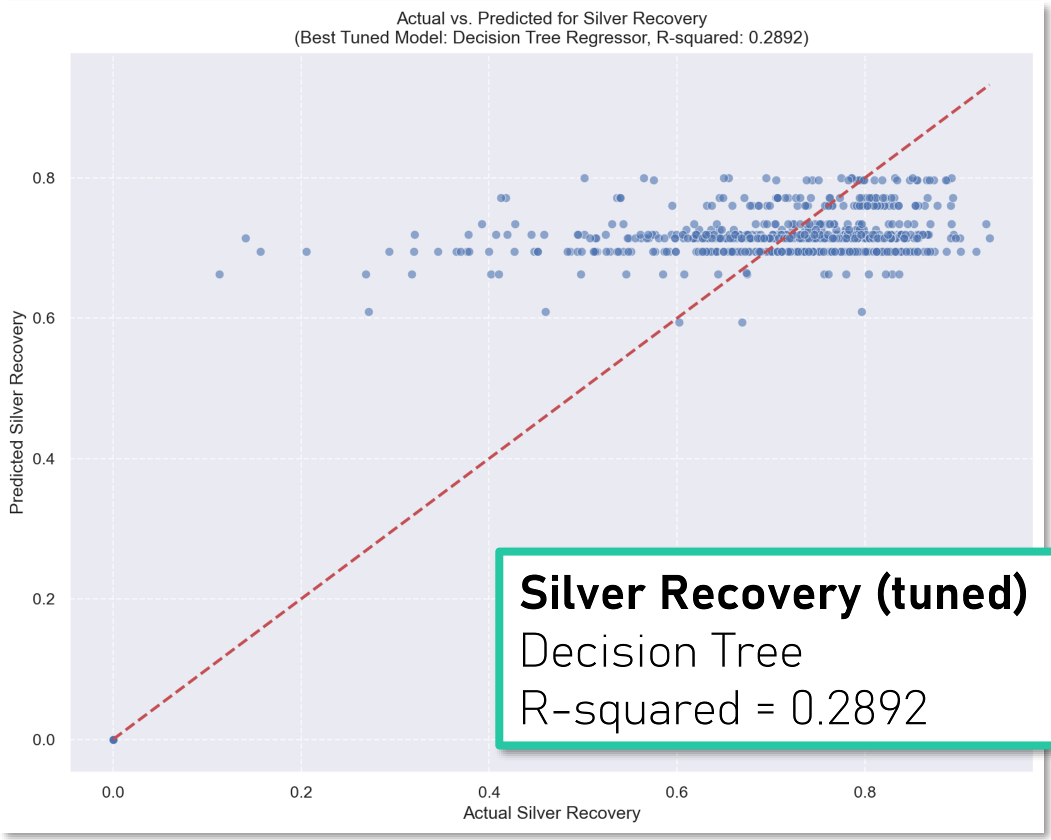

**Gold Ounces (tuned)**
Artificial Neural Network
R-squared = 0.9851

Using only Tonnage and Au, Ag grades, our model can explain 98.5% of the variation of our Gold Output. We are also able to predict with an error of ±8.956 oz Au, which is a relatively small error.

# What did we learn? – Silver Recovery



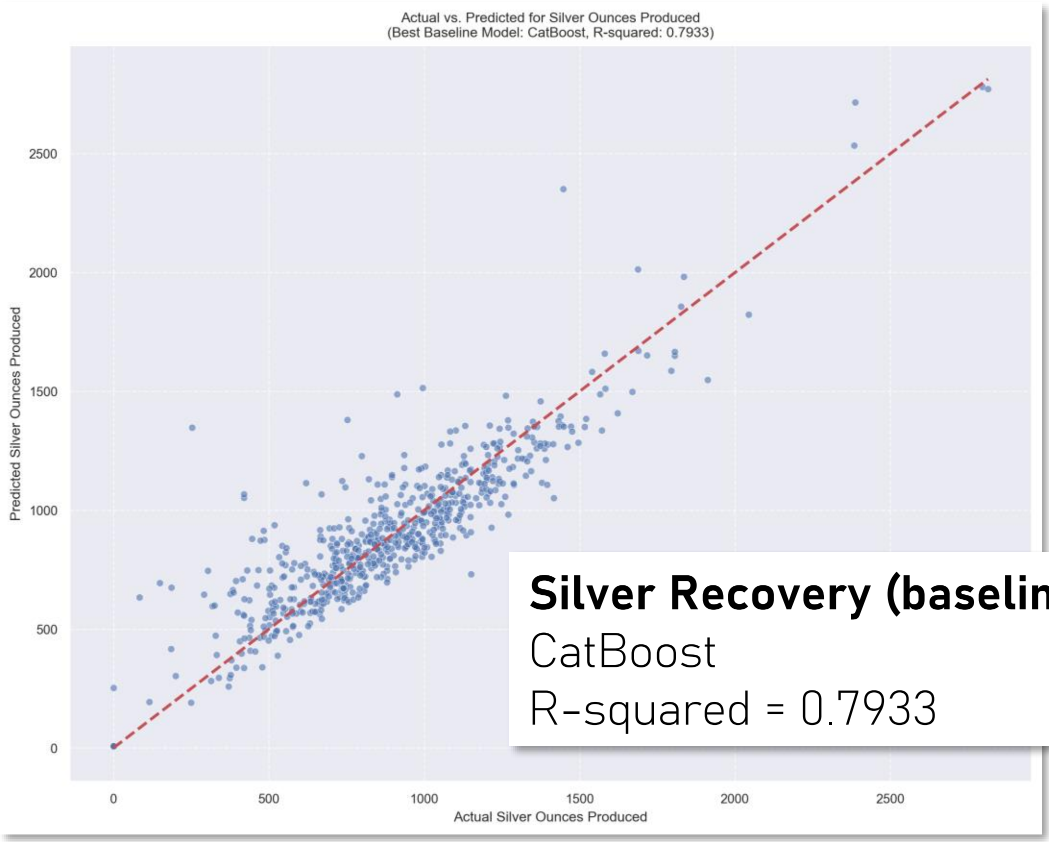

**Silver Recovery (baseline)**
Artificial Neural Network
R-squared = 0.2634



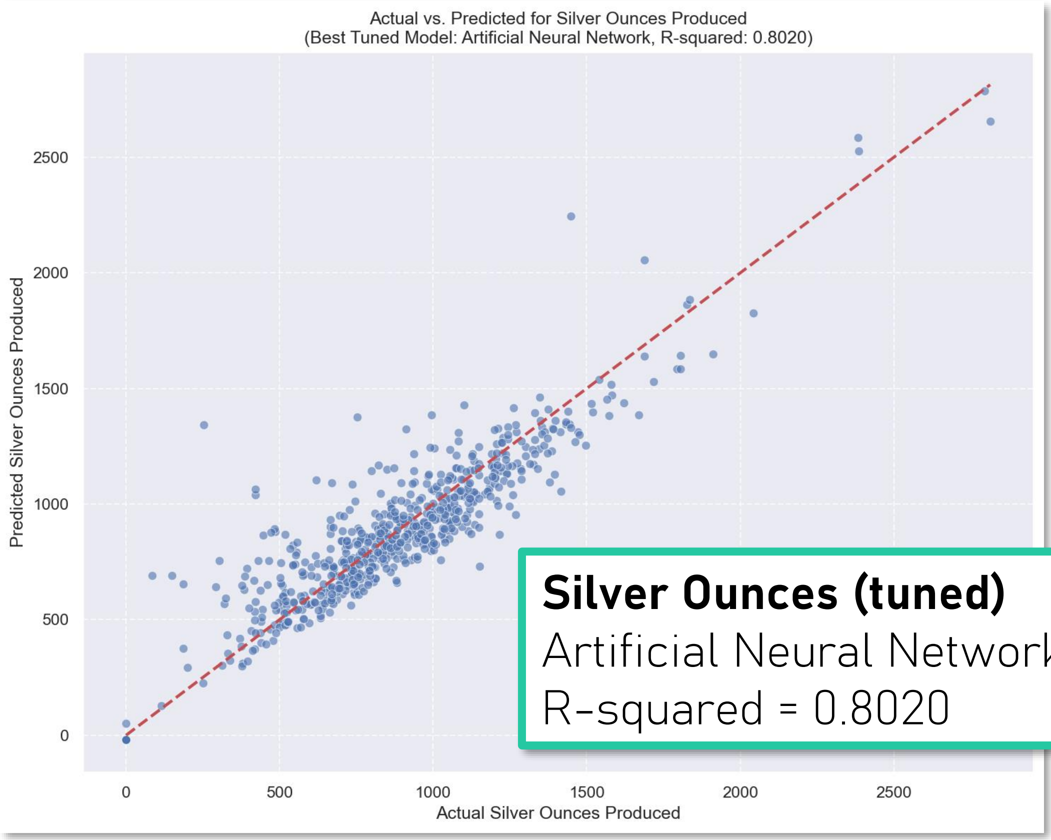

**Silver Recovery (tuned)**
Decision Tree
R-squared = 0.2892

Using only Tonnage and Au, Ag grades, our model can explain 28.9% of the variation of our Silver Recovery. We are also able to predict with an error of ±11.73%, which is quite a big error band. This means, we need other variables to better predict Silver Recovery.

# What did we learn? – Silver Ounces



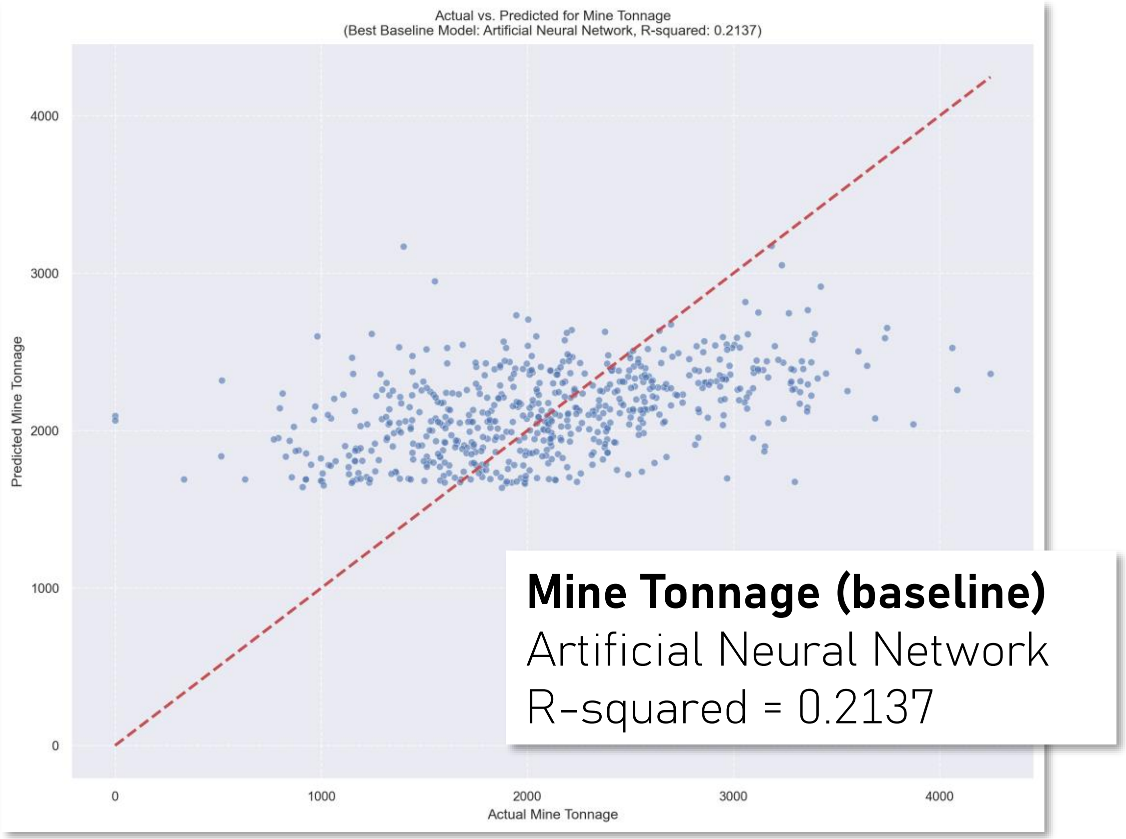

**Silver Recovery (baseline)**
CatBoost
R-squared = 0.7933

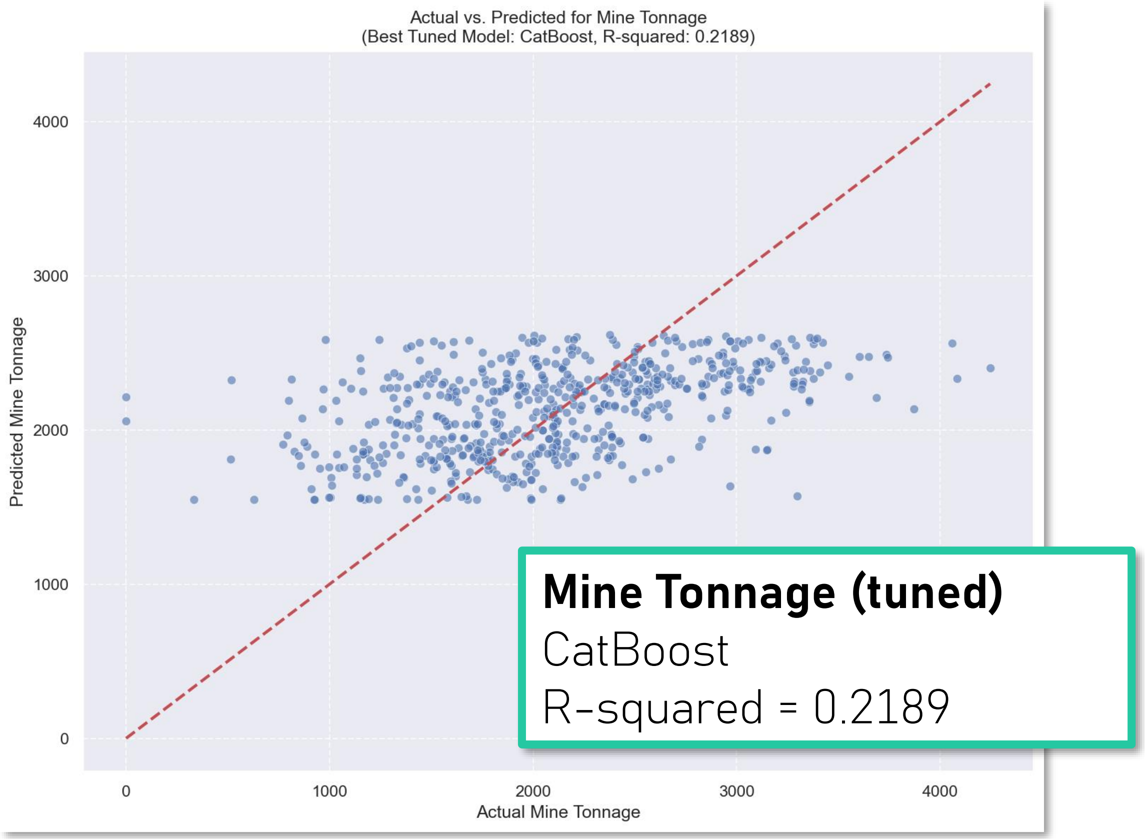



**Silver Ounces (tuned)**
Artificial Neural Network
R-squared = 0.8020

Using only Tonnage and Au, Ag grades, our model can explain 80.2% of the variation of our Silver Output. We are also able to predict with an error of ±151.45 oz Ag. Adding variables may further improve this model's accuracy.

# What did we learn? – Mine Tonnage

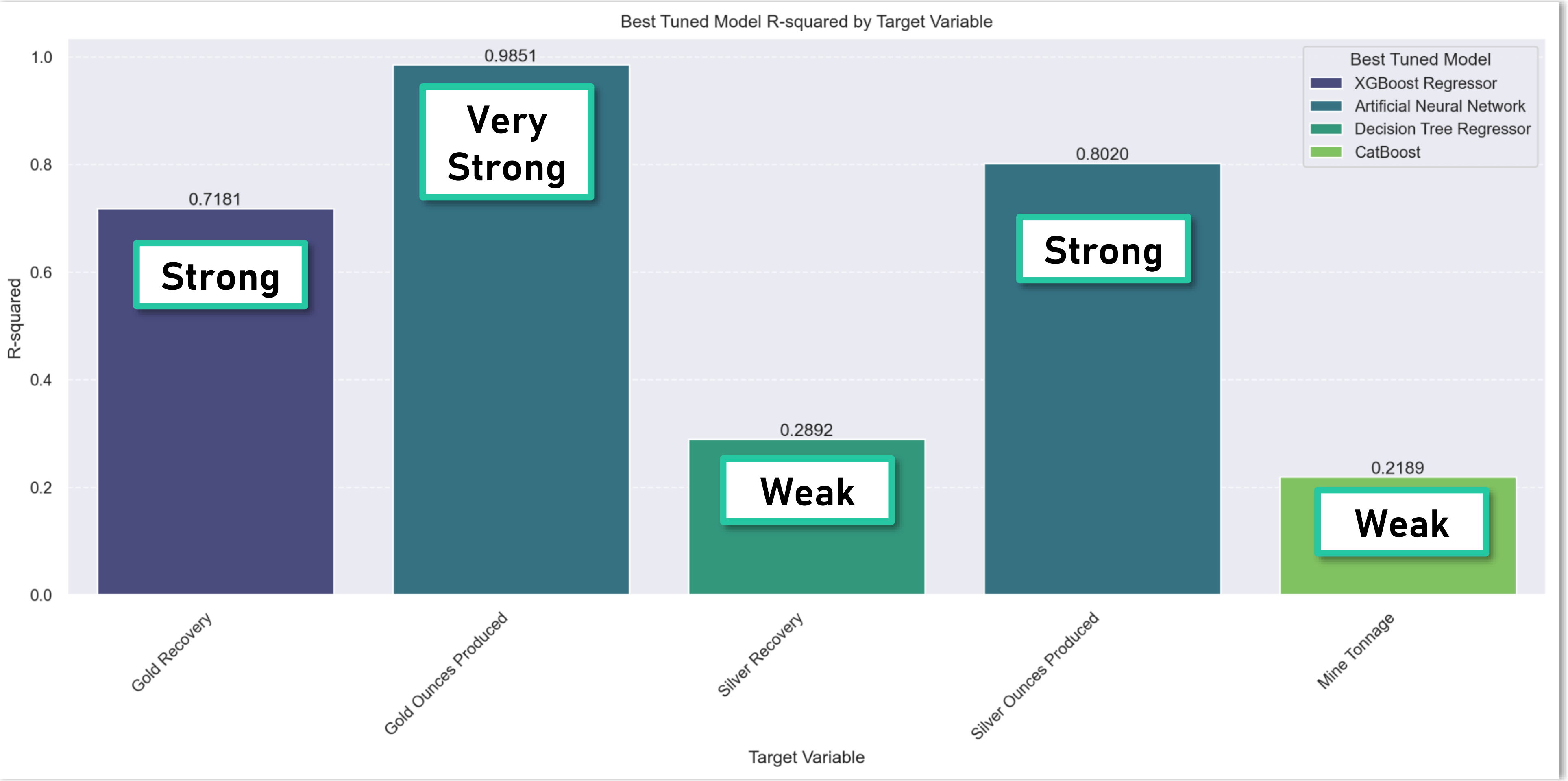



Using only Tonnage and Au, Ag grades, our model can explain 21.9% of the variation of our Mine Tonnage. We are also able to predict with an error of ±594.19 tons, which is a relatively poor performance. Adding variables may further improve this model's accuracy.

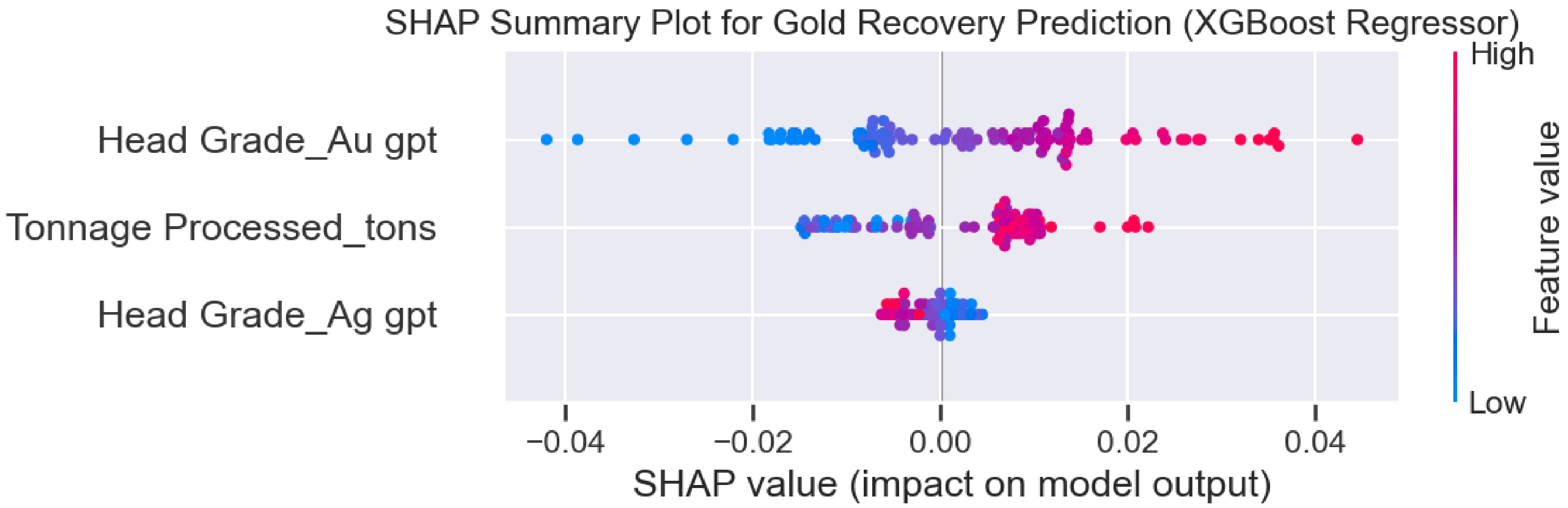
# Final Model Selection

| Target | Regression Model | R-squared | RMSE |
|---|---|---|---|
| Gold Ounces Produced | Artificial Neural Network (Multi Layer Perceptron) | **0.9851** | **8.9568** |
| Gold Recovery | XGBoost Regressor | **0.7181** | **0.0472** |
| Silver Ounces Produced | Artificial Neural Network | **0.8020** | **151.4564** |
| Silver Recovery | Decision Tree Regressor | **0.2892** | **0.1173** |
| Mine Tonnage | CatBoost | **0.2189** | **594.1986** |



# What affects our Gold Recovery?

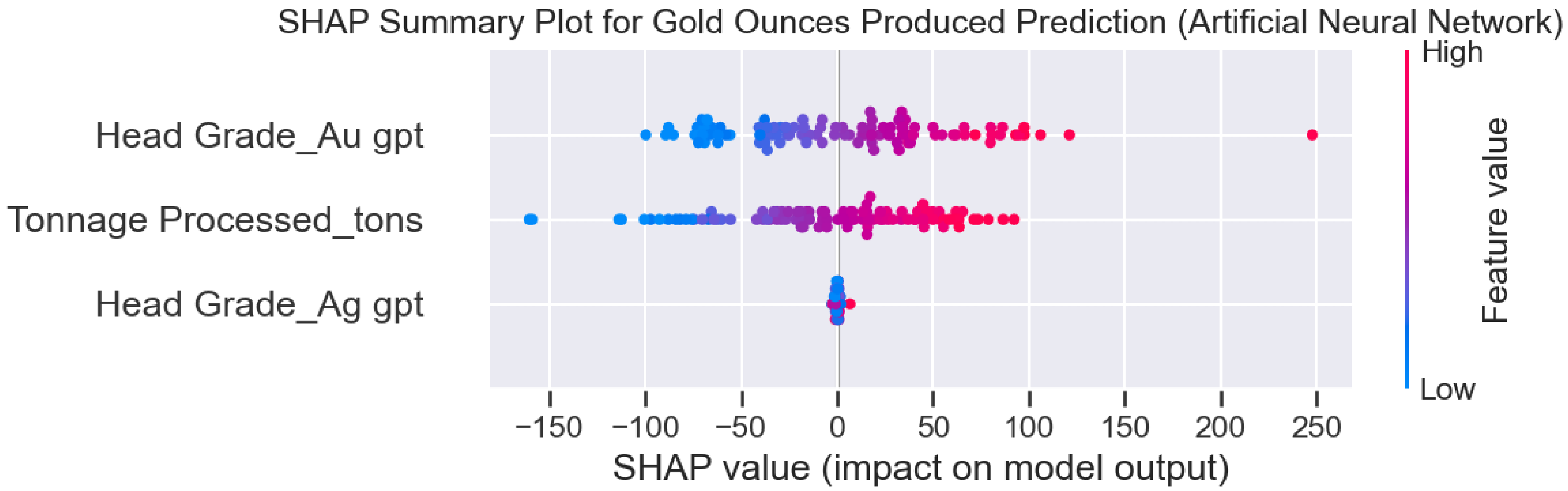
**Gold Recovery (tuned)**
XGBoost
R-squared = 0.7181




- **Most Influential Feature:** 'Head Grade_Au gpt' appears to be the most influential feature
- Higher 'Head Grade_Au gpt' generally leads to higher gold recovery.

# What affects our Gold Output?

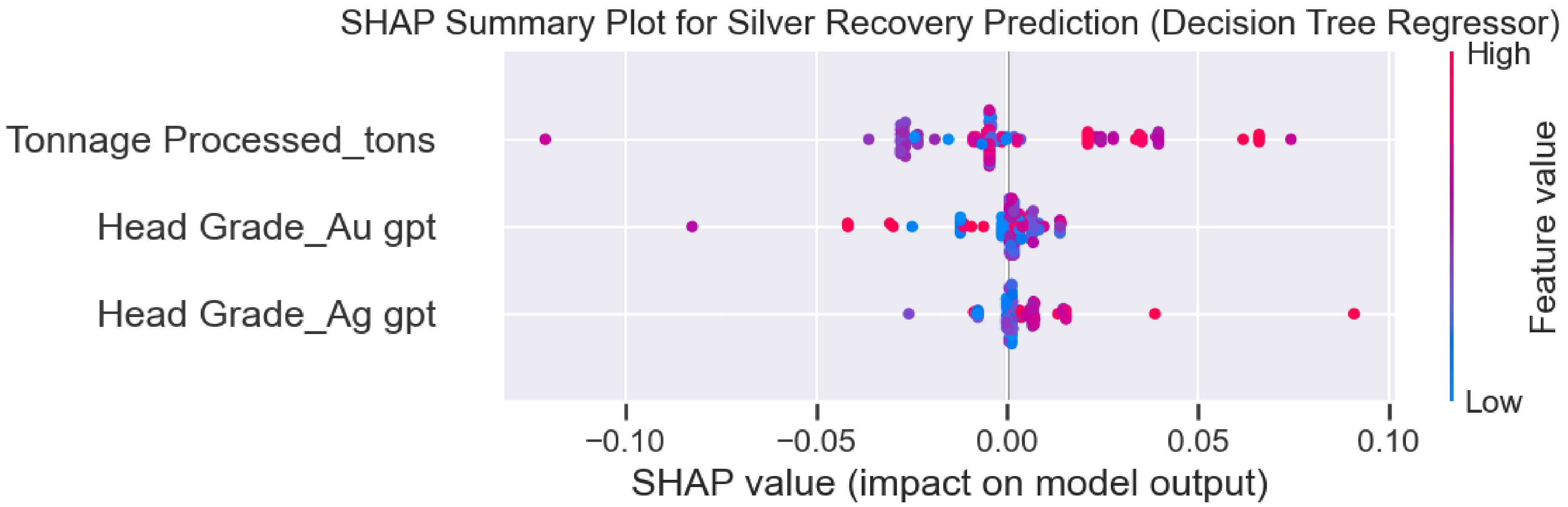
**Gold Ounces (tuned)**
Artificial Neural Network
R-squared = 0.9851




- **Most Influential Feature:** All three features ('Head Grade_Au gpt', 'Tonnage Processed_tons', 'Head Grade_Ag gpt') are highly important.
- 'Head Grade_Au gpt' and 'Tonnage Processed_tons' show a strong positive correlation with the output, meaning higher values of these features lead to higher predicted gold ounces produced.

# What affects our Silver Recovery?

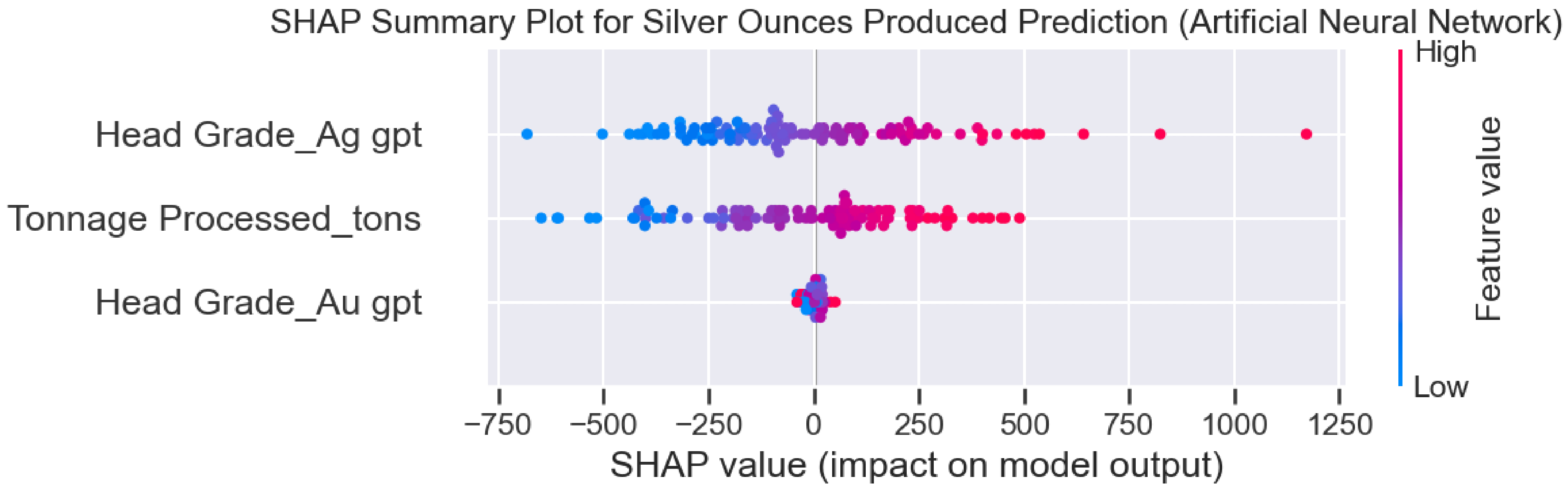
**Silver Recovery (tuned)**
Decision Tree
R-squared = 0.2892




- **Most Influential Feature:** 'Head Grade_Ag gpt' is the most impactful feature, which is intuitive. 'Head Grade_Au gpt' and 'Tonnage Processed_tons' have less impact.
- Deviations in 'Head Grade_Ag gpt' primarily influence the predicted silver recovery. Given the lower R-squared, the model's reliance on 'Head Grade_Ag gpt' suggests that other uncaptured factors likely play a significant role.

# What affects our Silver Output?

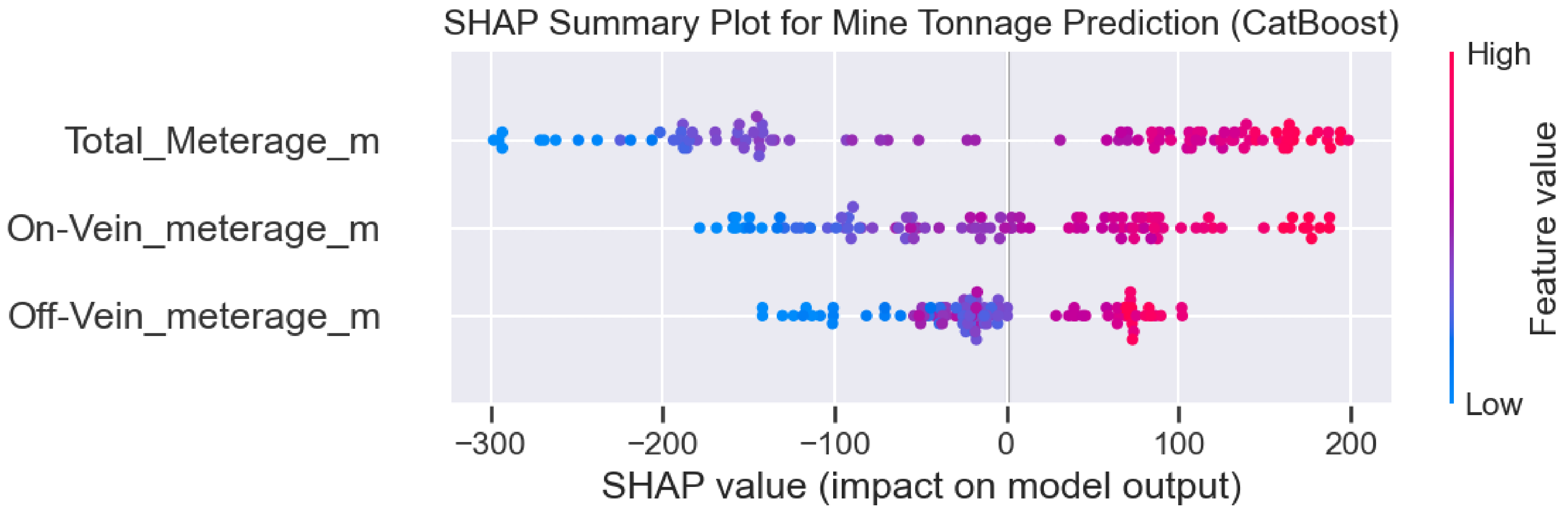
**Silver Ounces (tuned)**
Artificial Neural Network
R-squared = 0.8020




- **Most Influential Feature:** Similar to gold ounces, all three features contribute significantly.
- 'Head Grade_Ag gpt' and 'Tonnage Processed_tons' are particularly dominant, with higher values leading to increased silver ounces produced.

# What affects our Mine Tonnage?

**Mine Tonnage (tuned)**
CatBoost
R-squared = 0.2189




- **Most Influential Feature:** 'Total_Meterage_m' and 'Off-Vein_meterage_m' appear to be the most important features. However, the overall low R-squared for this target variable suggests that even these features have limited predictive power.
- 'Total_Meterage_m' contribute the often modest push or pull on the prediction. This reaffirms the earlier observation that the current feature set is insufficient for accurate mine tonnage prediction.

# Conclusion and Recommendation: Key Findings and Best Models

- **Gold Ounces Produced:**
  - **Best Model (Tuned):** Artificial Neural Network (R-squared: **0.9851**, RMSE: 8.9568)
  - This target variable is highly predictable, indicating strong relationships with the input features. Both ANNs and ensemble methods (CatBoost, XGBoost) achieved excellent performance.
  - **SHAP Insights:** 'Head Grade_Au gpt' and 'Tonnage Processed_tons' are highly important, showing a strong positive correlation with the output.
  - Out of all the target variables, gold ounces can be predicted most accurately with a RMSE of ±8.9568 oz Au which is low enough error for prediction and can already be used reliably for forecasting purposes.

# Conclusion and Recommendation: Key Findings and Best Models

- **Gold Recovery**:
  - **Best Model (Tuned):** XGBoost Regressor (R-squared: **0.7181**, RMSE: 0.0472)
  - Good predictive power was achieved. Ensemble methods (XGBoost, CatBoost, LightGBM) consistently outperformed traditional models.
  - **SHAP Insights:** 'Head Grade_Au gpt' is the most influential feature, positively impacting gold recovery, followed by 'Tonnage Processed_tons' and 'Head Grade_Ag gpt'.
  - Gold recovery can be predicted relatively accurate with a RMSE of ±0.0472% which can still be improved for higher prediction but can already be used reliably as guidance when needed.

# Conclusion and Recommendation: Key Findings and Best Models

- **Silver Ounces Produced**:
  - **Best Model (Tuned):** Artificial Neural Network (R-squared: **0.8020**, RMSE: 151.7891)
  - Moderate predictability was observed, better than silver recovery but not as reliable as gold ounces.
  - **SHAP Insights:** 'Head Grade_Ag gpt' and 'Tonnage Processed_tons' are dominant features, leading to increased silver ounces produced when their values are higher.
  - Silver ounces performed can be predicted accurate enough with a RMSE of ±151.7891 oz Ag which can still be improved for higher predictive power but can already be used reliably as guidance when needed.

# Conclusion and Recommendation: Key Findings and Best Models

- **Silver Recovery:**
  - **Best Model (Tuned):** Decision Tree Regressor (R-squared: **0.2892**, RMSE: 0.1173)
  - This proved to be the most challenging prediction task. Even after tuning, R-squared values remained low, suggesting the current feature set explains only a small portion of its variance. Traditional models like Linear Regression and Decision Trees performed surprisingly well after tuning compared to some ensemble methods for this specific target.
  - **SHAP Insights:** 'Head Grade_Ag gpt' is the most impactful feature. However, the low R-squared suggests other uncaptured factors play a significant role.
  - Silver Recovery performed not as well as the other target variables only achieving an RMSE of ±0.1173% which can still be improved for higher predictive power but can already be used reliably as guidance when needed
  - A deeper investigation into additional, more relevant features is crucial. This could involve consulting domain experts to identify factors like geological formations, ore body characteristics, equipment operational parameters, or temporal trends that might influence these metrics.

# Conclusion and Recommendation: Key Findings and Best Models

- **Mine Tonnage**:
  - **Best Model (Tuned):** Support Vector Machine (R-squared: **0.1995,** RMSE: 588.6655)
  - All models struggled significantly with this prediction. The R-squared values were very low (around 0.20), indicating that the selected predictors (`Off-Vein_meterage_m`, `On-Vein_meterage_m`, `Total_Meterage_m`) explain very little of the variance in 'Mine Tonnage_tons'.
  - **Therefore it is recommended that the current available feature set be expanded to other mine production metrics: such as blast width, stoping height, etc..**
  - **SHAP Insights:** 'Total_Meterage_m' and 'Off-Vein_meterage_m' were identified as important, but the overall low R-squared indicates limited predictive power from the current feature set for this target.
  - Mine Tonnage performed the worst out of all the models achieving an RMSE of ±588.66 tons which is too high of an error to predict it accurately. This only means the model can still be improved a lot by adding other features.
  - A deeper investigation into additional, more relevant features is crucial. This could involve consulting domain experts to identify factors like other mining metrics, equipment operational parameters, or temporal trends that might influence these metrics

# Conclusion and Recommendation: Major Data Gaps

- **Time Series Analysis**
  - Time-Series analysis was not considered due to the **broken datasets especially during COVID times** wherein proper data-recording was not done at that time.
  - Time-Series analysis was also considered unsuitable due to the tendency of ARIMA and LSTM models to possibly overshoot projections because it is **not bound by limiting factors** such as number of Mining Equipment, Milling Equipment, Mechanical Availability, etc.
- **Mine Production and Mine Advance Data**
  - Majority of data available are **not synced at the same time-period as the other daily data in the mill**, but it can possibly be summarized or generalized to fit the same time frame as the data used in this study.
- **Data Granularity**
  - As much as some datasets are recorded per-shift, some where not recorded with any time or date stamps therefore they are much harder to index and incorporate in this dataset.
  - Data granularity was set to daily since it is the **most common report interval between the mine and the mill**. A more detailed granularity of this report will probably increase model predictive performance especially for the lower performing models in this capstone project.

# Conclusion and Recommendation: Major Insights Gained

- The predictive power of a Machine Learning model is dependent on **the strength of the relationships** between the predictor variables with respect to the target variable.
- A **good balance** between model accuracy and the **required predictor variables** are required to come up with a decently accurate prediction.
  - i.e. A model with 10 inputs may give out a better or more accurate prediction than one requiring only 2 to 3 inputs, but having to record 10 inputs will be **much more cumbersome or possibly too expensive and time consuming** as to the weight or impact of the prediction to a company's process.
- The **timeliness of a prediction** must also be considered. If a prediction can only be made too close to when the official results are obtained, decreases the necessity for the prediction since it doesn't give ample time for the company to adapt or change its parameters to improve the outcome of the target variable.
  - This also means that the predictor variables used to forecast **the predictor variable should already have been recorded much earlier** in the process for it to have any meaningful impact to the company.
  - Therefore, a timely prediction for % Recovery will only be meaningful if the recorded input metrics or features **occur prior** to when the target metal/s have already been completely extracted from the ore.

# Conclusion and Recommendation: Items for Consideration

- **Feature Engineering for Challenging Predictions:**
  - For 'Silver Recovery' and especially 'Mine Tonnage', a deeper investigation into additional, more relevant features is crucial. This could involve consulting domain experts to identify factors like geological formations, ore body characteristics, equipment operational parameters, or temporal trends that might influence these metrics.
- **Explore Time-Series Models:**
  - Given that all datasets are time-indexed, exploring time-series specific models (e.g., ARIMA, Prophet, LSTMs for sequential data) could capture temporal dependencies that traditional regression models might miss, especially for recovery rates and mine tonnage.
- **Advanced Hyperparameter Optimization:**
  - While `GridSearchCV` was used, other advanced optimization techniques like `RandomizedSearchCV` with broader distributions or Bayesian Optimization may be employed for a more exhaustive search of the hyperparameter space, especially for ANNs and complex ensemble models

# Conclusion and Recommendation: Items for Consideration

- **Ensemble Stacking/Blending:**
  - Combining predictions from several best-performing models (e.g., CatBoost, XGBoost, ANN) through stacking or blending could potentially yield even better performance for each target variable.
- **Further SHAP Analysis:**
  - Investigate SHAP dependence plots for critical features to understand how a feature's value affects the prediction when interacting with other features. This could provide deeper insights into the underlying process dynamics.
- **Data Quality and Granularity:**
  - Assess if more granular or higher-quality data points are available, particularly for the challenging predictions. The low R-squared values for 'Mine Tonnage' might suggest that the current features are too high-level or miss crucial information that will improve model accuracy.


# Thank You.

Jose Norbiel G. Florendo

AIM-PGDAIML

josenorbielflorendo@gmail.com