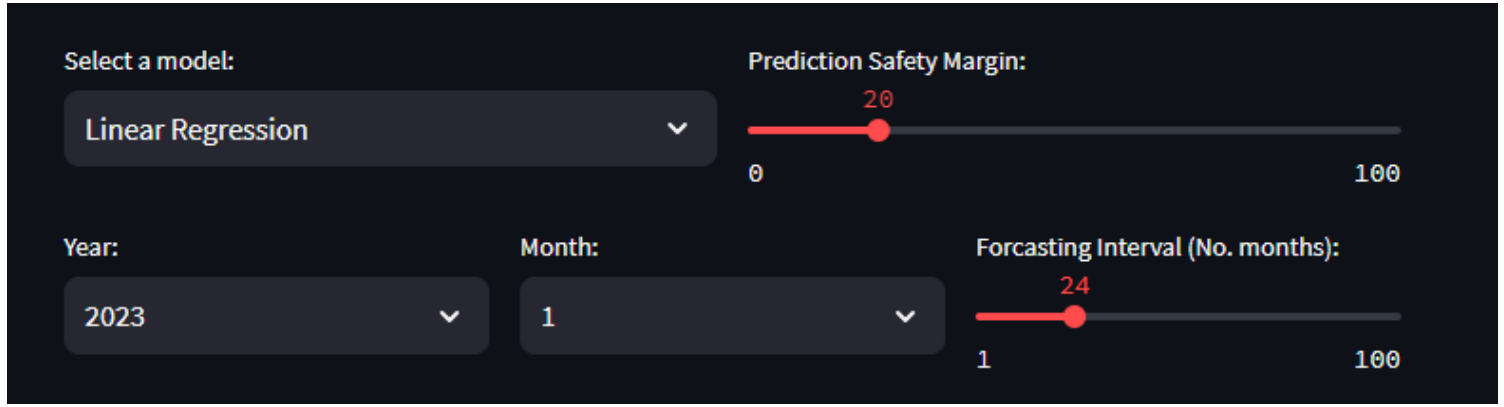
Case Study Data Scientist Production Planning

Anahita Pakiman

Consumption forecast



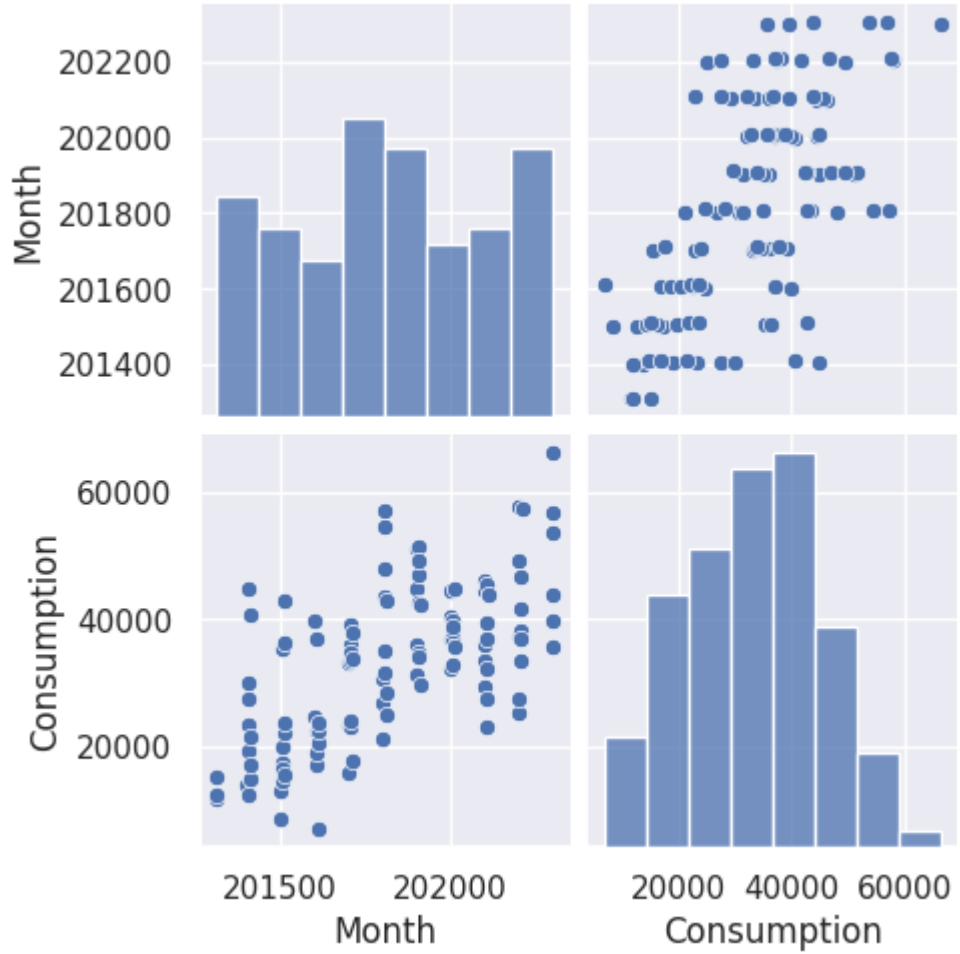
Predict Consumption APP

Models:

- Linear Regression - annual
- Random Forest - monthly
- Decision tree - monthly

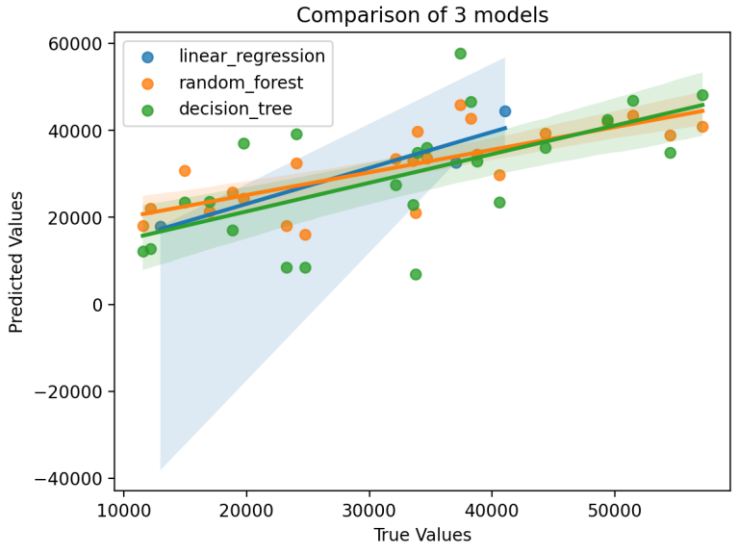
$$Consumption = pred + SM * RMSE$$

Data Cleaning



Models

Predict Consumption APP

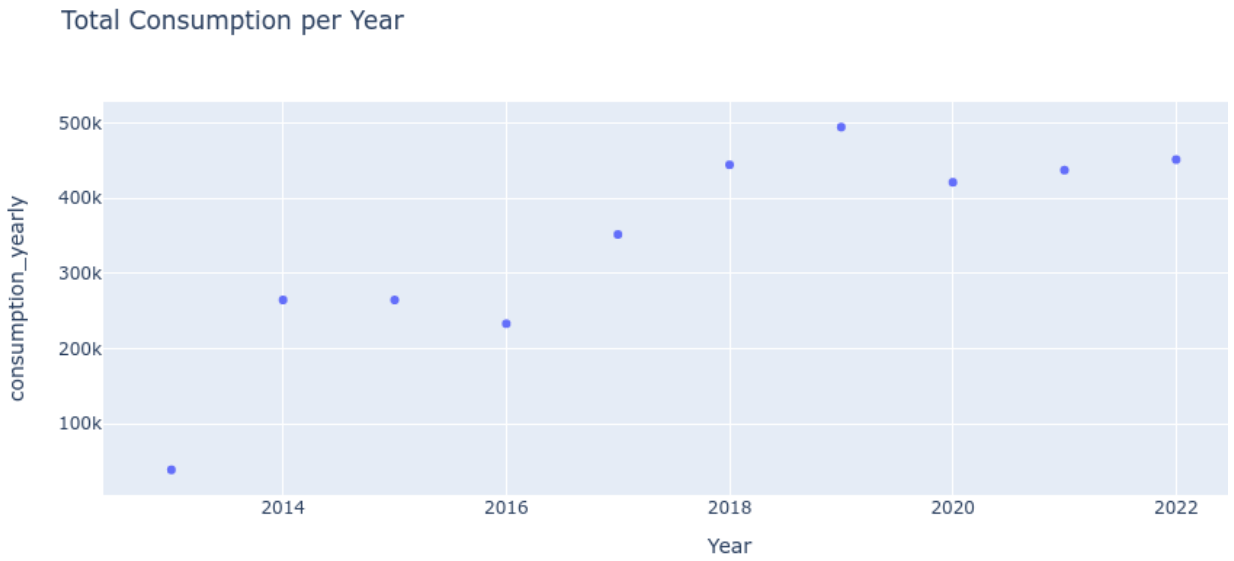


Methods	RMSE
• Mean	10262
• Linear Regression	4295
• Random Forest	8486
• Decision Tree	11642

Linear Regression:
- RMSE on the test data: 4295
- The future cost prediction: 99622
- with 20% safety margin = 100481

Random Forest:
- RMSE on the test data: 8486
- The future cost prediction: 1186575
- with 20% safety margin = 1188272

- Decision Tree: - RMSE on the test data: 11642
- The future cost prediction: 1274226
- with 20% safety margin = 1276554



Improvement

- **Data split**:
  Change data split for annual forecast, first split and later average annual or seasonal consumption.

- **Hyperparameter tuning**
  All models are set to their default hyperparameters. Of the models examined, the decision tree clearly overfits the data and the model complexity needs to be reduced (max_depth).

- **Modify prediction target, Time Series**
  Current models are trained to predict for a specific year/month. However, the result could be improved if we use a time series prediction and consider the time history (e.g. average consumption) in the input data and train it on its future prediction performance.

- **Re-evaluate data cleaning**
  Check the validity of the cleaned data points and reconsider including them in the dataset.

- **Try other models**
  For example: Support Vector Regression (SVR), Autoregressive Integrated Moving Average (ARIMA)

Commercial Plant Manager

- Present the APP and the parameters study available in the APP.
- Highlight the current best prediction and the reasoning behind it.
- Ask him/her to use the app and give feedback on the visualisations he/she needs to improve the use case and understanding of the models.
- Suggest further selection of data for better prediction (e.g. material type).

Predict Consumption APP