**Neural Network Documentation- XPERTSIM**

1. Overview

This document provides an in-depth explanation of the two neural networks used to predict economic indicators:

* Neural Network1 for rig\_count, tam and capex. It depends on both year and country.
* Neural Network2 for gas\_price and oil\_price. It is year specific network which is having same values for countries.

Each model has distinct characteristics, input features, and training methodologies.

Each value prediction depends on specific indicators from the global and industrial datasets, as outlined below:

**Gas Price Prediction:** Uses Oil Price, Rig Count, Infrastructure Index, Capex

**Oil Price Prediction:** Uses Rig Count, Infrastructure Index, Capex, Political Stability.

**Rig Count Prediction:** Uses Oil Price, Gas Price, Capex, Infrastructure Index, GDP Growth, Political Stability.

**Tam Prediction:** Uses Capex, GDP, Population, Health Index, Education, Infrastructure Index.

**Capex Prediction:** Oil Price, Gas Price, Rig Count, GDP Growth, Political Stability, Infrastructure Index.

2. Data Processing & Preprocessing

Before training, the datasets undergo the following preprocessing steps:

**Loading Data**

* The data is imported from CSV or Excel files and stored in a database across multiple tables.
* The function load\_data(dataset\_name) loads the data into a Pandas DataFrame.
* The country\_name column is converted to uppercase for consistency.
* The year column is ensured to be an integer.

**Handling Missing Data**

* Data is sorted by year.
* Missing values are filled using linear interpolation.

Feature Scaling

* The MinMaxScaler normalizes features to a 0 to 1 range.
* Separate scalers are used for input (X) and target (y).

3. Neural Network 1

These models predict rig count, tam, and capex for each country individually.

**Architecture**

* Input Layer: Multiple input features.
* Hidden Layer 1:

Dense(64, activation='relu') (64 neurons, ReLU activation)

BatchNormalization() (stabilizes activations)

Dropout(0.2) (prevents overfitting)

* Hidden Layer 2:

Dense(32, activation='relu') (32 neurons)

BatchNormalization(), Dropout(0.2)

* Output Layer:

Dense(1, activation='linear') (predicts one numerical value)

Training Process

* Train-Test Split: 80% training, 20% testing.
* Optimizer: Adam (adaptive learning rate adjustment).
* Loss Function: Mean Squared Error (mse).
* Callbacks:

ReduceLROnPlateau: Lowers learning rate when loss stagnates.

EarlyStopping: Stops training if no improvement.

Making Predictions

* The latest available data is used as the base.
* A random trend factor (0.9 to 1.1) is applied to simulate economic fluctuations.
* Predictions are rounded for rig\_count as it must be an integer.

Influence on Predictions

✔ Captures country-specific trends.

✔ Uses recent data for accurate forecasting.

✖ Limited historical data may cause poor generalization.

✖ Random trend factor introduces some noise.

4. Neural Network 2

These models predict gas and oil prices, assuming a global economic influence rather than country-specific factors.

**Architecture**

* Input Layer: Only year is used as input.
* Hidden Layers:

Dense(64, activation='relu')

BatchNormalization(), Dropout(0.2)

Dense(32, activation='relu')

* Output Layer:

Dense(1, activation='linear') (predicts price)

Training Process

* Data Grouping: Data is grouped by year, averaging all values.
* Feature Scaling: year and target are scaled separately.
* Train-Test Split: 80% training, 20% testing.
* Optimizer & Loss Function: Same as country-specific models.

Making Predictions

* Future years are transformed into the scaled range.
* The model predicts scaled prices, which are later inverted back.

Influence on Predictions

✔ Captures macroeconomic trends.

✔ Efficient single-model prediction.

5. Prediction Integration

Once models are trained, predictions are merged into a final dataset.

**Process Overview**

* A skeleton DataFrame (skeleton) is created with:
* All selected countries.
* All years from start\_year to end\_year.

**Gas & Oil Price Predictions:**

Single model predicts values for all selected countries.

**Other three Predictions:**

Separate models predict rig\_count, tam, and capex per country.

The predictions are merged into a csv file and then stored in database.