**Supervised Learning- Time Series Use Case with Daily Female Birth Prediction**

**Problem**

* **What problem are you solving?**

Predicting the number of female births on a given day.

* **Why is it worth solving?**

Accurate predictions can help in planning and resource allocation in healthcare and related sectors.

* **What is the source of your data and what kind of data are you using?**

The dataset used is the "Daily Female Births” containing the number of daily female births.

**Data Collection**

This dataset offers details on a range of daily female birth:

Variables:

Date: various date format: (YYYY-MM-DD)

Births: Number of births

****

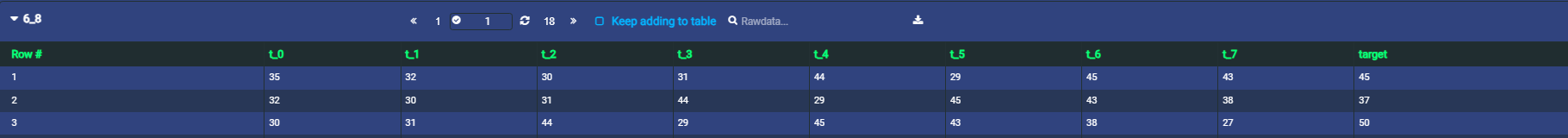
**Methodology**

**Data Wrangling (Time Lag & Window):**

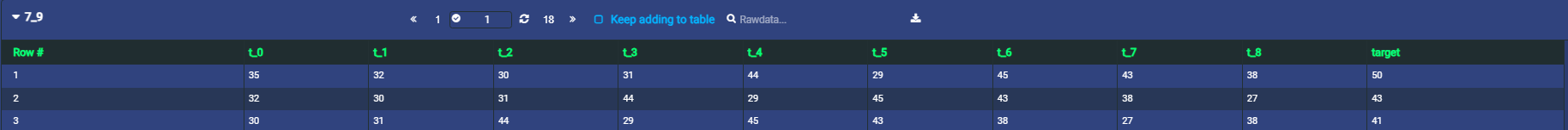
* **1 Time Lag (Forecast Time) & 3 Window Size (Historical Value):**

****

* **6 Time Lag (Forecast Time) & 8 Window Size (Historical Value):**

****

* **7 Time Lag (Forecast Time) & 9 Window Size (Historical Value):**



**Univariate Analysis**

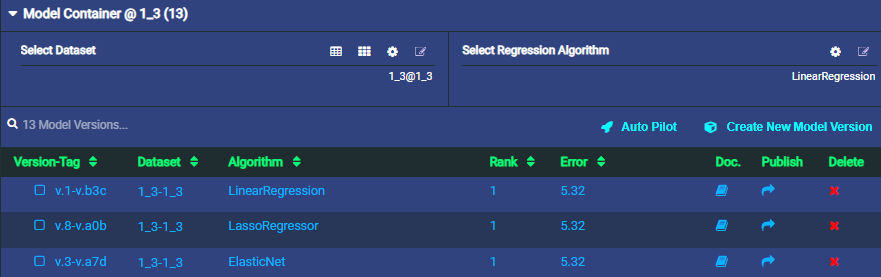
* Analysed the date column (e.g., months, days) to identify patterns or trends.
* Evaluated data distribution, missing values, and outliers.

**Data Preparation**

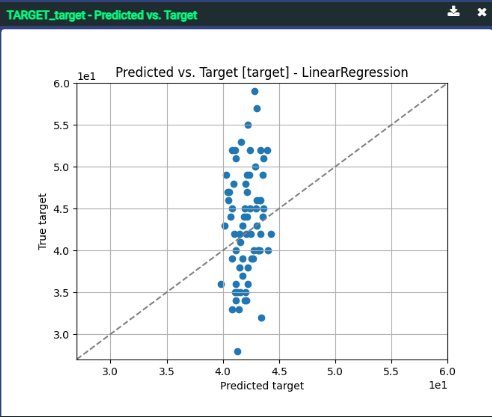
* Extracted relevant features from the date column, such as day of the week, month, and any significant dates.
* Handled missing values and outliers appropriately.
* Normalized/standardized the data if necessary.

**Time Lag & Window:**

**1 Time Lag (Forecast Time) & 3 Window Size (Historical Value):**

****

**Target vs Predicted**

****

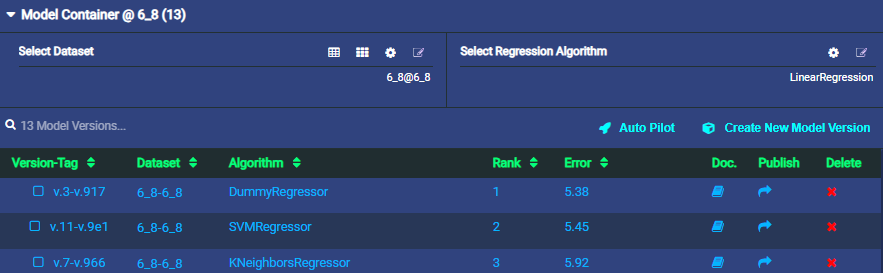
**Evaluation Metrics**

Used metrics such as Mean Squared Log Error, Median Absolute Error, Mean Absolute Error (MAE), and Mean Squared Error (MSE).

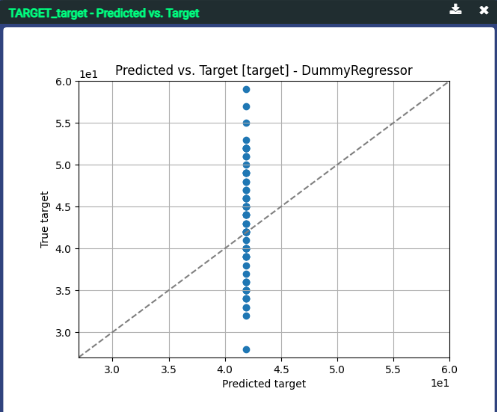
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression** | **Mean Squared Log Error** | **MAE** | **MSE** | **Median Absolute Error** |
| LinearRegression | 0.02 | 5.32 | 6.45 | 4.75 |
| LassoRegressor | 0.02 | 5.31 | 6.46 | 4.79 |
| ElasticNet | 0.02 | 5.31 | 6.46 | 4.77 |

This window size 1 and 3 shows the top three algorithms with the lowest error rate, and the target / predicted graph refers to the dots that are close to the median line. Each algorithm establishes evaluation metrics mean and median absolute, squared log, and error.

**6 Time Lag (Forecast Time) & 8 Window Size (Historical Value):**

****

**Target vs Predicted**

****

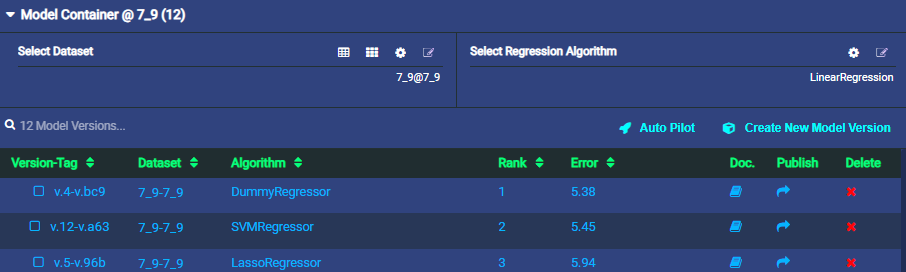
**Evaluation Metrics**

Used metrics such as Mean Squared Log Error, Median Absolute Error, Mean Absolute Error (MAE), and Mean Squared Error (MSE).

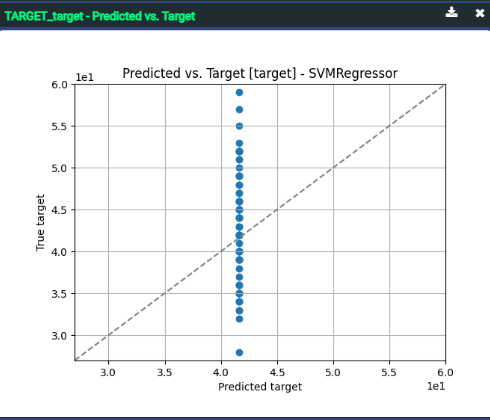
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression** | **Mean Squared Log Error** | **MAE** | **MSE** | **Median Absolute Error** |
| DummyRegressor | 0.02 | 5.38 | 6.64 | 4.91 |
| SVMRegressor | 0.02 | 5.44 | 6.70 | 4.59 |
| KNeighbors | 0.02 | 5.92 | 7.11 | 5 |

This window size 6 and 8 shows the top three algorithms with the lowest error rate, and the target / predicted graph refers to the dots that are close to the median line. Each algorithm establishes evaluation metrics mean and median absolute, squared log, and error.

**7 Time Lag (Forecast Time) & 9 Window Size (Historical Value):**

****

**Target vs Predicted**

****

**Evaluation Metrics**

Used metrics such as Mean Squared Log Error, Median Absolute Error, Mean Absolute Error (MAE), and Mean Squared Error (MSE).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression** | **Mean Squared Log Error** | **MAE** | **MSE** | **Median Absolute Error** |
| DummyRegressor | 0.02 | 5.37 | 6.64 | 4.92 |
| SVMRegressor | 0.02 | 5.44 | 6.70 | 4.59 |
| LassoRegressor | 0.02 | 5.93 | 7.23 | 4.87 |

This window size 7 and 9 shows the top three algorithms with the lowest error rate, and the target / predicted graph refers to the dots that are close to the median line. Each algorithm establishes evaluation metrics mean and median absolute, squared log, and error.

**Modelling Process**

**Algorithm Selection**

Used regression algorithms like Linear Regression, Elastic Net, SVM Regressor, Dummy Regressor, Lasso Regressor, K Neighbors Regressor

**Linear Regression:**

Linear regression uses the relationship between the data-points to draw a straight line through all them. This line can be used to predict future values.

**SVM Regressor:**

Creates a Linear SVR object using the Vertica SVM (Support Vector Machine) algorithm. This algorithm finds the hyperplane used to approximate distribution of the data.

**Elastic Net:**

Elastic net linear regression uses the penalties from both the lasso and ridge techniques to regularize regression models

**Dummy Regressor:**

This regressor is useful as a simple baseline to compare with other (real) regressors. Do not use it for real problems

**Lasso Regressor:**

Lasso regression—also known as L1 regularization—is a form of regularization for linear regression models

**K Neighbors Regressor:**

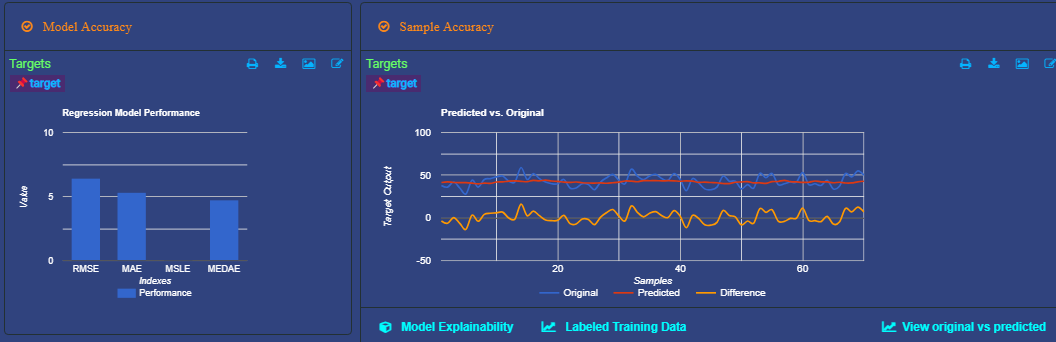
The target is predicted by local interpolation of the targets associated of the nearest neighbors in the training set

**Best Result (Linear Regression)**

Linear Regression is the best model in a window frame of one time lag (forecast time) and three window sizes (historical value)

The error rate is 5.32, and the graph's dots are close to the median

**Model Performance**

****

**Performance Metrics**

****

**Feature Importance**

****

**Result Summary**

Presented the results of the best-performing model Linear Regression with

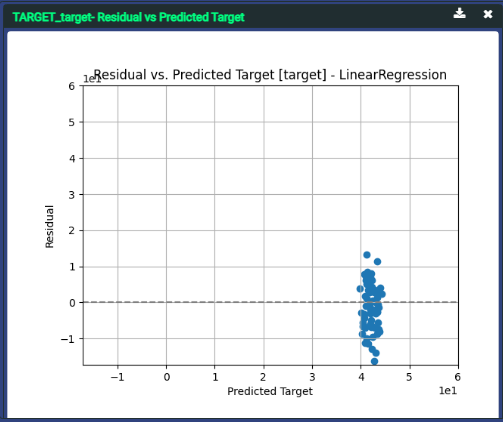
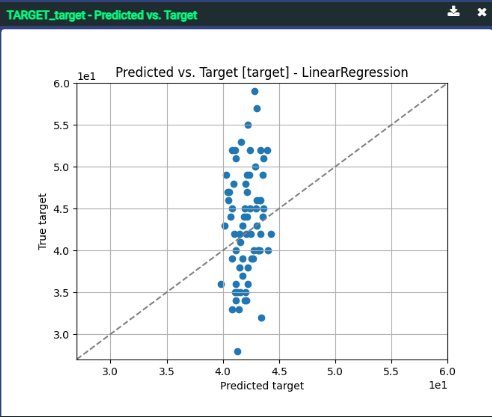
Mean Squared Log Error: 0.02162

MAE: 5.32119

MSE: 6.45583

Median Absolute Error: 4.75796

**Target vs Predicted**

**** ****

**Conclusions**

1. **Future Improvements**
   * Discussed potential improvements like incorporating more features, using more advanced models, or refining the current model.
2. **Real-life Application**
   * Explained how this predictive model could be used in real-life scenarios, such as hospital staffing or resource planning.
3. **Value to Client**
   * Highlighted the value this solution could bring to stakeholders, such as improved planning and resource allocation.
4. **Learnings**
   * Summarized the key learnings and insights gained from the project.