

ABSTRACT

Forecasting time-series data is a common practice these days, and with various methods available, such as statistical methods or machine learning approaches. The main goal of this research is to propose a hybrid of two machine learning models which are Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM) for sales revenue forecasting. Which then will be compared to a Vanilla LSTM and ARIMA model.

This research will analyze the time-series dataset from a Kaggle Competition held in 2015. The dataset consists of Rossmann Store Sales split into an 80-20 train test data. The data will be preprocessed into a Min-Max Scaler then reshaped into a 4-dimensional input for the CNN model to extract features then fed into the LSTM model for predictions based on the feature extraction.

The result of implementing these three methods gave different results in terms of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) scoring for the evaluation method, with the CNN-LSTM model giving the best score of the evaluation test in RMSE and MAE. Although the MAPE score is lower, due to the sensitivity of the model this number concludes that feature engineering is crucial for the model.

Keywords: *Revenue, Time-series, Forecasting, CNN-LSTM, LSTM, ARIMA, Deep Learning, RMSE*