Case study: univariate time series analysis and forecasting of pharmaceutical products' sales data at small scale

Milan Zdravković*, Jasmina Đorđević**, Aleksandra Catić-Đorđević***, Sonja Pavlović****, Miodrag Ivković******

* Faculty of Mechanical Engineering, University of Niš, Niš, Serbia
** Faculty of Science and Mathematics, University of Niš, Niš, Serbia
** Department of Pharmacy, University of Niš, Niš, Serbia
** Neven Farma, Niš, Serbia
** Technical Faculty "Mihajlo Pupin", University of Novi Sad, Novi Sad, Serbia
milan.zdravkovic@gmail.com, djordjevichristina@gmail.com, aleksandra1610@yahoo.com, sonjapav@gmail.com, misa.ivkovic@gmail.com

***Abstract*— The objective of the research behind the paper was to validate different methods and approaches related to sales time series data preparation, analysis and forecasting, with aim to facilitate recommending sales and marketing strategies based on trend/seasonality effects and forecasting sales of eight different groups of pharmaceutical products with diverse characteristics, such as stationarity, seasonality, amount of residuals and sales data variance. All these analyses and forecasts are made on a small scale, for a single distributor, pharmacy chain or even individual pharmacy. Paper presents only research work related to univariate time series analysis, while potential candidates for explanatory input variables were also identified and shortly elaborated. Effectiveness of three forecasting methods, namely ARIMA, Facebook's Prophet and Long-Short Term Memory (LSTM) neural networks was investigated. Each of the method is complemented with two optimization and validation approaches, relevant for short-term (so called rolling forecast scenario) and long-term forecasting.**

I. INTRODUCTION

On a larger scale, the sales forecasting in pharmaceutical industry [1] is typically done by using Naïve model, where the forecasted values equal values in the previous period with added factor of growth, which is specifically defined for different regions, markets, categories of products, etc. Although this model fails when the market saturates, in general and on a larger scale, it has proven as relatively successful.

Still, analysis and forecasts on a smaller scale, such as single distributor, pharmacy chain or even individual pharmacy, smaller periods such as weeks, etc., guide very important decisions related to resource and procurement planning, what-if analyses, return-on-investment forecasting, business planning and others.

The main problem in smaller scale time series analyses and forecasts are significant uncertainties and sales performance very close to random, making the forecasts with accuracies above thresholds as defined by Naïve methods difficult to achieve.

The main research question we tackle is related to exploring the feasibility of use of modern time-series forecasting methods in pharmaceutical products sales forecasting on a smaller scale. In specific, we benchmark the accuracies achieved with those methods against the performances of basic Naïve, Seasonal Naïve and Average methods.

This paper is structured into 3 main parts. First, research methodology, actually a problem-neutral time series forecasting pipeline is presented. Next, the actual implementation is presented, by highlighting the steps made in following the proposed methodology in the case of pharmaceutical products sales data analysis and forecasting. Finally, the conclusion brings the final interpretation of actual results and some suggestions to the sales department, driven by the result of the data analysis.

II. METHODOLOGY

The methodology for implementing this case study follows the typical time series forecasting pipeline, consisting of three major phases: 1) feature engineering and data preparation; 2) exploratory data analysis (time-series analysis); and 3) forecasting.

Based on the problem and objective formal definition, the data acquired from the sales information system are cleaned, feature engineering approach was defined, and all data are transformed to hourly time series, consisting of aggregate sales among different classes of pharmaceutical products in hourly time periods, namely: anti-inflammatory and antirheumatic products (M01AB, M01AE), analgesics and antipyretics (N02BA, N02BE), psycholeptics drugs (N05B, N05C), drugs for obstructive airway diseases (R03) and antihistamines for systemic use (R06). This, intermediary series is used for the formal definition of anomalies and their identification. Also, outliers are detected in consultation with pharmacy staff and treated by first, imputing the missing data and then, by imputing representative data, by using several methods. Finally, data is then rescaled to weekly time-series and stored [2].

Time series analysis had two-fold objective. First, annual, weekly and daily data analysis were done with objective to make potentially useful conclusions and propositions for improving sales and marketing strategies. Then, stationarity, autocorrelation and predictability analysis of the time series in individual groups was analyzed to infer the initial set of parameters for implementing the forecasting methods.

Forecasting was carried out at the weekly scale. Two different approaches to forecasting problem scoping were adopted. First one implements so called rolling forecast, namely forecasting the sales in the next week, by using the model trained with all historical data. Therefore, during testing, prediction in a timestep t is based on the model which fits the training set consisting of observations in timesteps (0,t-1), or: f(t) = f(o[0:t-1]). Rolling forecast model can be used for short-term resource planning and planning the procurement of stock of pharmaceutical products.

Another approach is related to long-term forecasting, for example forecasting the future period of one year, by using the model trained with historical data. This model can be used for business planning and making decisions of strategic nature. Train-test split validation with one last year of data (52 rows) was used for testing. Key performance indicator for forecasting accuracies in both approaches was Mean Squared Error (MSE). Baseline accuracy was calculated by using Naïve and Seasonal Naïve, for rolling forecasts and Average method for long-term ones.

Three different models were tested: ARIMA/SARIMA [3] (for rolling and long-term forecast), Facebook's Prophet [4] (for rolling and long-term forecast) and Long-Short Term Memory (LSTM) [5] artificial neural network architectures (for long-term forecast).

Grid search was used as adopted approach for hyper-parameters optimization, for ARIMA and Prophet model. LSTM was applied only for long-term forecasting.

The data preparation process for LSTM included transforming to stationary time series, sequencing time series to $[X_{t-n\_steps}...X_{t-2},X_{t-1}][y_t]$ shape (previously determining input vector dimension which gives best accuracies), transforming to supervised learning problem and time series scaling (normalization). Three LSTM architectures were tested: Vanilla LSTM, Stacked LSTM and Bidirectional LSTM.

A. *Forecasting models*

ARIMA (Auto-Regressive Integrated Moving Average) [3] models are most commonly used tools for forecasting univariate stationary time-series. Model uses the dependency relationship (correlation) between an observation and some number of lagged observations (AR) in the past. It is integrated (I), namely it uses differencing (see above) to make time-series stationary, within the method. Finally, it uses the dependency between an observation and a residual error from a moving average model applied to lagged observations (MA).

Prophet [4] is Facebook's additive regression model, that includes: linear or logistic trend, yearly seasonal component modeled using Fourier series and user-provided list of important holidays. The model facilitates easy customization and reliable forecasts with default configurations. According to the authors, Prophet is successful for forecasting data with strong "human-scale" seasonality (day of week, time of year), reasonable number of missing data and/or outliers, historical trend changes, non-linear trends (with saturation), at least one year of observations, known holidays.

Long-Short Term Memory (LSTM) [5] are a form of Recurrent Neural Networks (RNN) - deep learning architectures that are characterized by the use of LSTM units in hidden layers. Main feature of RNNs is that they allow information to persist, or they can inform the decision on some classification or regression task in the moment t, by using observations (or decisions) at moments t-1, t-2,.., t-n. In this research, three different LSTM architectures were used. Vanilla LSTM [5] is made of a single hidden layer of LSTM units, and an output layer used to make a prediction. Stacked LSTM is architecture with two or more hidden layers of LSTM units stacked one on top of another. In bidirectional LSTM architecture model learns the input sequences both forward and backward.

III. SOLUTION AND DISCUSSION

Initial dataset consisted of 600000 transactional data collected in 6 years (period 2014-2019), indicating date and time of sale, pharmaceutical drug brand name and sold quantity. As a result of the interviews with pharmacists, decision was made that the subject of analyses and forecasting will be actual drug categories, instead of the individual drugs.

For the further processing, transformation of the dataset is being done, to the structure indicated on the Figure 1. Transformation included: creating columns with different ATC codes, inserting data corresponding to the sale of the particular drug for the given time, multiplied with a quantity and some data cleaning. Selected group of drugs (57 drugs) is classified to 8 Anatomical Therapeutic Chemical (ATC) Classification System categories: M01AB, M01AE, N02BA, N02BE, N05C, N05B, R03 and R06. ATC codes features are added to the dataset and data was resampled to the hourly time-series.

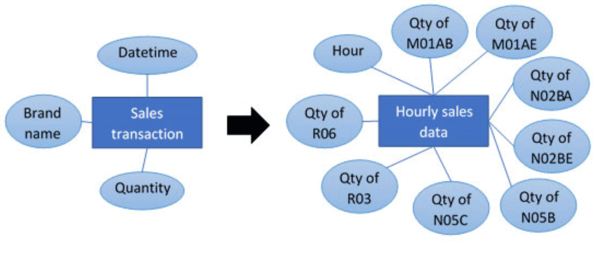

Figure 1. Transformation of the initial dataset

Anomalies and outlier data detection and treatment was carried out next, based on data in hourly time series. Initial analysis of the data has shown rare individual anomalies in the form of spikes of sales for all drug categories in particular hours, indicating human error in data entry or irrelevant entry.

The outliers are assumed as unexpected (by the univariate forecasting model) spikes or drops of sales - those data points lying far outside of the interval of confidence for the forecasting model trained with all data in the time series, where the distance was determined as a function of value and upper limit of the interval of confidence: D = (value-yhat_upper/value) for sales spikes (no drops were visible in the time series).

All data points with D>0.8 (for N02BE D>0.75, for N05C D>0.9) are then treated as outliers and replaced with NaN, which was also done for anomalies. Finally, for individual missing data imputation, mean of seasonal counter-parts method was used. For imputation of continual missing data, forecasting missing data by using Prophet approach was chosen. Prepared data was resampled to weekly series and stored.

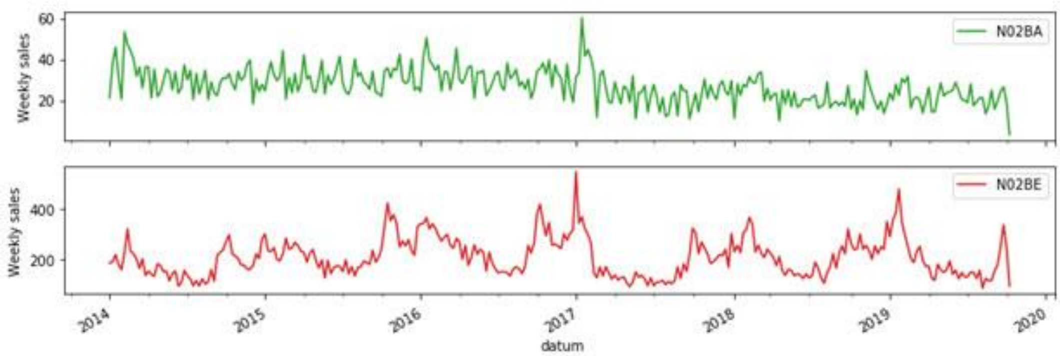

Figure 1. Weekly sales data for N02BA and N02BE categories, after outlier and anomalies treatment

Time Series Analysis included stationarity, autocorrelation, seasonality, regularity and data distribution analysis. Augmented Dickey-Fuller (ADF) test have shown that all data, but N02BA (P-value=0.249) in the series were stationary, with maximum confidence. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test found the trend non-stationarity in N02BE, R03 and R06. Moderate autocorrelation is observed at ACF (Auto-Correlation Function) and PACF (Partial Auto-Correlation Function) plots for all series, with exception of N05C sales. N02BE, R03 and R06 series were found to exhibit annual seasonality. For calculating regularity and predictability of time series, Approximate Entropy test was used. For all series, entropy values were higher than 1 indicating low predictability, with highest values for M01AE, M01AB and N02BA.

Forecasting models were fitted with weekly time-series data with dataset of 302 rows. Three forecasting methods were tested: ARIMA, Prophet and LSTM. Train-test split method was used (52 weeks of test data). MSE was used as metrics for comparing the performance and also as a loss function for LSTM. In order to define the baseline forecasting accuracy to improve from, three tests were performed. Average method was used as a baseline for long-term forecasting, while Naïve and Seasonal Naïve were used for rolling forecasts. Table 1 shows MSE for rolling forecast accuracy for different models, with included Naïve models for benchmarking.

Method/Drug ATC category	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06
ARIMA	71.56	76.57	31.94	2614.39	149.91	7.98	668.13	75.06
Auto-ARIMA	76.87	80.35	33.51	2147.07	147.13	8.02	666.68	74.35
Prophet	88.62	81.61	38.67	3010.05	208.84	9.14	823.13	81.33
Naïve	116.01	93.88	44.74	2753.64	255.49	14.92	948.56	82.23
Seasonal Naïve	N/A	N/A	N/A	2923.37	N/A	N/A	1131.4	87.29

Table 1. MSE for different forecasting methods applied on time series with drug categories sales data – rolling forecast

For rolling forecast, ARIMA method (Auto-ARIMA for series with seasonal character) outperforms Prophet and is considered as a best candidate for short-term sales forecasting.

Table 2 shows the MSE of the long-term forecasting accuracy for different LSTM model configurations, compared also with Auto-ARIMA and Prophet performance in long-term forecasting. Displayed results for LSTM were achieved with default configurations of hyper-parameters, as indicated at the beginning of this section. In general, as expected, LSTM models have shown better performance with largest uptake of what is considered as randomness, for example in N05B. In other cases, competitive performance was achieved, even without the hyper parameters' optimization.

Method/Drug ATC category	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06
Vanilla LSTM	85.38	91.43	39.12	2585.91	174.82	9.23	687.72	71.78
Stacked LSTM	84.71	116.5	35.38	2683.48	184.69	9.16	701.85	66.84
Bi-directional LSTM	90.8	103.05	34.55	2777.74	198.02	9.54	1016.76	68.62
Facebook Prophet	69.62	79.58	47.35	3089.93	305.56	10.2	825.42	76.74
Auto ARIMA	77.64	99.07	51.2	3850.61	178.21	7.99	1169.78	111.38
Average	75.43	99.09	76.46	6837.27	162.00	8.05	1191.19	186.82

Table 2. MSE for different forecasting methods applied on time series with drug categories sales data – long-term forecasting

IV. CONCLUSION

To conclude, time-series analyses and forecasts have guided potentially useful conclusions and recommendations to the pharmacy. Daily, weekly and annual seasonality analysis were proven useful for identifying the periods in which special sales and marketing campaigns could be implemented, except for N05B and N05C categories of drugs which did not exhibit significant regularities. Forecasts have proven better than Naïve methods and in acceptable intervals for long-term planning. It is highly likely that the forecasts could be significantly improved by expanding the problem scope to multivariate time series forecasting and by including explanatory variables, such as:

- Weather data. Sales of antirheumatic drugs in M01AB and M01AE categories could be affected by the changes of atmospheric pressure. Sudden declines in all categories could be explained by extreme weather conditions, such as heavy rain, thunderstorms and blizzards.
- Price of the drugs. Sales spikes may be explained by the discounts, applied in a short term. Introducing this feature may facilitate what-if forecasting analysis of sales performance during marketing campaigns involving price reductions.
- Dates of the pension payoff. Sales spikes are visible at the dates of state pensions payoff.
- National holidays, as non-working days with seasonal patterns similar to Sundays are expected to disrupt daily sales.

Future work on univariate time series forecasting includes increasing the number of data, exploring different other accuracy metrics, optimization of hyper-parameters for LSTM models and testing other architectures, such as CNN LSTM and ConvLSTM. However, key improvements in sales forecasting are expected from reducing the uncertainty of the models by expanding to multivariate time series forecasting problem, as explained above.

ACKNOWLEDGMENT

This research was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Cook, A.G. (2015) Forecasting for the Pharmaceutical Industry: Models for New Product and In-Market Forecasting and How to Use Them. Routledge. ISBN-13: 978-1472460110

[2] Zdravković, M. Pharma sales data. Kaggle, doi: 10.34740/KAGGLE/DS/466126

[3] Arunraj, N.S., Ahrens, D., Fernandes, M. (2016) Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. International Journal of Operations Research and Information Systems (IJORIS) 7(2):21

[4] Taylor, S.J., Letham, B. (2017) Forecasting at scale. Preprint. https://peerj.com/preprints/3190/

[5] Hochreiter, S., Schmidhuber, J. (1997) Long Short-Term Memory. Neural Computation. 9(8):1735-1780

.