MEF UNIVERSITY

ELECTRICITY DEMAND FORECASTING FOR TURKEY

Capstone Project

Hakan Yiğit

İSTANBUL, 2018

MEF UNIVERSITY

ELECTRICITY DEMAND FORECASTING FOR TURKEY

Capstone Project

Hakan Yiğit

Advisor: Prof. Semra Ağralı

İSTANBUL, 2018

MEF UNIVERSITY

Name of the project: Electricity Demand Forecasting for Turkey
Name/Last Name of the Student: Hakan Yiğit
Date of Thesis Defense: 03/09/2018

I hereby state that the graduation project prepared by Hakan Yiğit has been completed under my supervision. I accept this work as a "Graduation Project".

03/09/2018
Prof. Semra Ağralı

I hereby state that I have examined this graduation project by Hakan Yiğit which is accepted by his supervisor. This work is acceptable as a graduation project and the student is eligible to take the graduation project examination.

03/09/2018
Prof. Özgür Özlük

Director
of
Big Data Analytics Program

We hereby state that we have held the graduation examination of Hakan Yiğit and agree that the student has satisfied all requirements.

THE EXAMINATION COMMITTEE

Committee Member	Signature
1. Prof. Semra Ağralı	………………………..
2. Prof. Özgür Özlük	………………………..

Academic Honesty Pledge

I promise not to collaborate with anyone, not to seek or accept any outside help, and not to give any help to others.

I understand that all resources in print or on the web must be explicitly cited.

In keeping with MEF University's ideals, I pledge that this work is my own and that I have neither given nor received inappropriate assistance in preparing it.

Hakan Yiğit	03.09.2018	Signature

EXECUTIVE SUMMARY

ELECTRICITY DEMAND FORECASTING FOR TURKEY

Hakan Yiğit

Advisor: Prof. Semra Ağralı

SEPTEMBER, 2018, 20 pages

Forecasting demand for products and services accurately provides competitive advantage to companies. This capstone project focuses on electricity demand forecasting for Turkey. Since energy storage is not yet a viable option, generated electricity is consumed simultaneously; and this fact exposes the importance of electricity demand forecast. Main aims of this project are to perform exploratory data analysis of the Turkish power market and apply machine learning algorithms to forecast electricity demand. Turkey's electricity demand is predicted using real-life data obtained for years between 2017 and 2018. The results show that electricity demand can be modeled using machine learning algorithms, and the models can be used to predict future electricity demand.

Key Words: Electricity demand forecasting, exploratory data analysis, machine learning.

ÖZET

TÜRKİYE’NİN ELEKTRİK TALEP TAHMİNİ

Hakan Yiğit

Prof. Dr. Semra Ağralı

EYLÜL, 2018, 20 sayfa

Ürünler ve hizmetler için doğru talep tahmini, rekabetçi avantajlar sağlayabilir. Bu proje, Türkiye'de elektrik talebinin tahmin edilmesine odaklanacaktır. Enerji depolama bugünlerde istenilen seviyede olmadığından, üretilen elektrik aynı anda tüketilmekte ve bu durum elektrik talebi tahminlerinin önemini ortaya koymaktadır. Projenin amacı Türkiye elektrik piyasasının veri analizini yapmak ve elektrik talebini tahmin etmek için makine öğrenimi algoritmalarını uygulamaktır. Türkiye'nin elektrik talebi 2017'den 2018'e kadar olan veriler kullanılarak tahmin edilmiştir. Sonuçlar, elektrik talebinin makine öğrenimi algoritmaları kullanılarak modellenebildiğini ve modellerin, gelecekteki elektrik talebini tahmin etmek için kullanılabileceğini göstermektedir.

Anahtar Kelimeler: Elektrik talep tahmini, data analizi, makine öğrenmesi.

TABLE OF CONTENTS

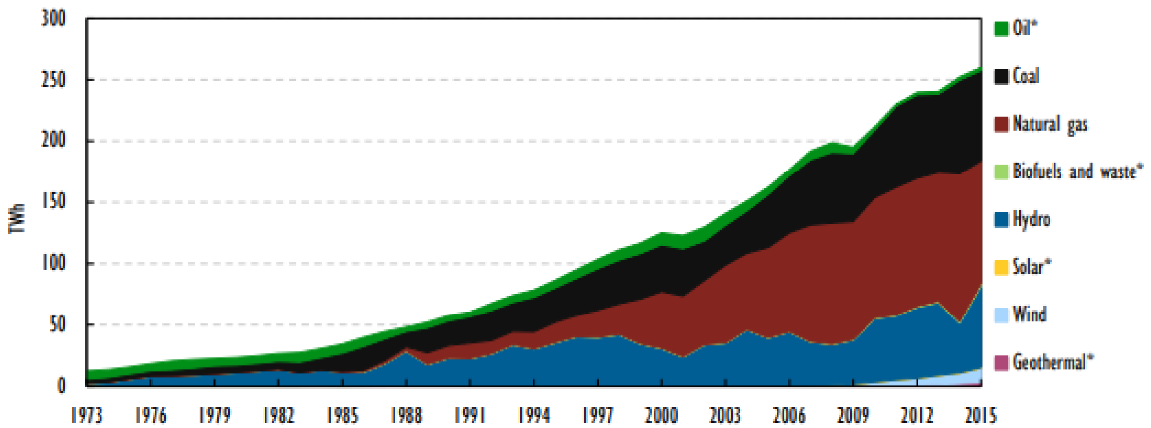
.

1. INTRODUCTION

Energy consumption is one of the most important indicators of the economic development for the countries. Especially, if there is a huge energy import dependency, as in Turkey, predicting future energy demand and supply becomes a vital topic. Turkey's energy import dependency is around 76% according to The Oxford Institute for Energy Studies (Gas Supply Changes in Turkey, 2018).

Over the past decade the Turkish economy has been growing at an average rate of 5%, which in turn has been a major driver in increasing energy demand and investment in the Turkish energy market. Over ten years (2004-14), electricity demand has been almost doubled and reached at 207 terawatt hours (TWh) in 2015 while gas demand has grown even faster, rising from 22 billion cubic meters (bcm^3) to 49 bcm^3. Since 2013-14, the economic growth has slowed down while the foreign debt, inflation and exchange rate volatility have increased. Turkey has a young and urbanizing population and expanding energy needs. Total primary energy supply (TPES) per capita and power generation per capita are still low in comparison to many other countries. However, TPES has risen considerably over the past 40 years from 24.4 Mtoe in 1973 to 129.7 Mtoe in 2015, and it is set to follow this trend in the coming decades. Electricity consumption grows at a pace comparable to the industrial boom phases of large economies (Energy Policies of IEA Countries, Turkey 2016).

Turkey's energy targets under the vision 2023 include the promotion of indigenous energy resources such as coal, having 30% share of renewable energy in electricity production, the reduction of energy intensity by 20% below 2010 levels through improved efficiency, and the launch of two nuclear power plants. Figure 1 shows the electricity generation by source between 1973 and 2015.

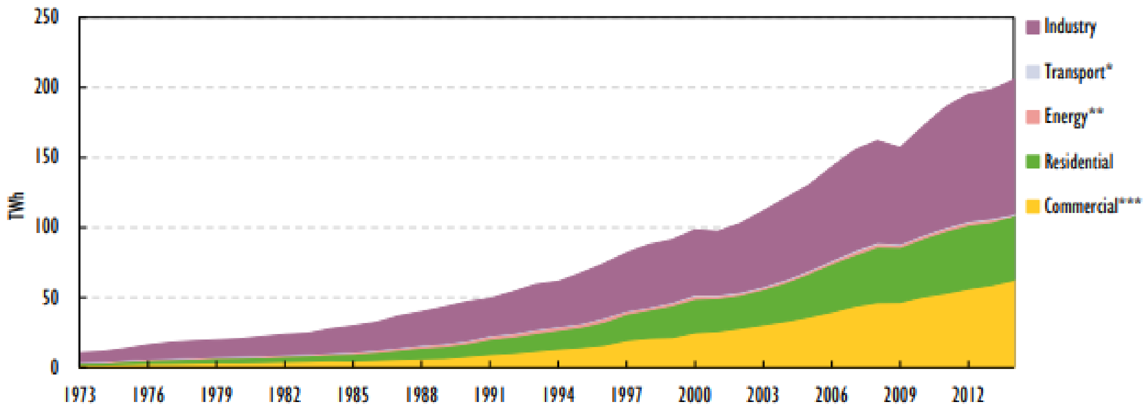


Note: Data are estimated for 2015.

* Negligible.

Figure 1. Electricity generation by source, 1973-2015, IEA (2016)

Figure 2 shows the electricity consumption of Turkey with respect to different sectors between 1973 and 2014. Industry has the largest share of electricity consumption with the rate of 46.2% of total consumption. Industry demand grew by 65% over the past decade. Commercial (includes commercial and public services, agriculture, fishing and forestry) and residential sector accounted for 30.1% and 22.3% of demand in 2014, respectively.



* Negligible.

** *Energy* includes coal mining, oil and gas extraction, and refining.

*** *Commercial* includes commercial and public services, agriculture, fishing and forestry.

Figure 2. Electricity consumption by sector, 1973-2014, IEA (2016)

According to the Electricity Demand Projection 2014-2035 (see Figure 3), Ministry of Energy and Natural Resources - MENR expects electricity demand to increase more than three times of the current level during the next 20 years. The industrialization, urbanization and population growth are major components of demand increase according to the report.

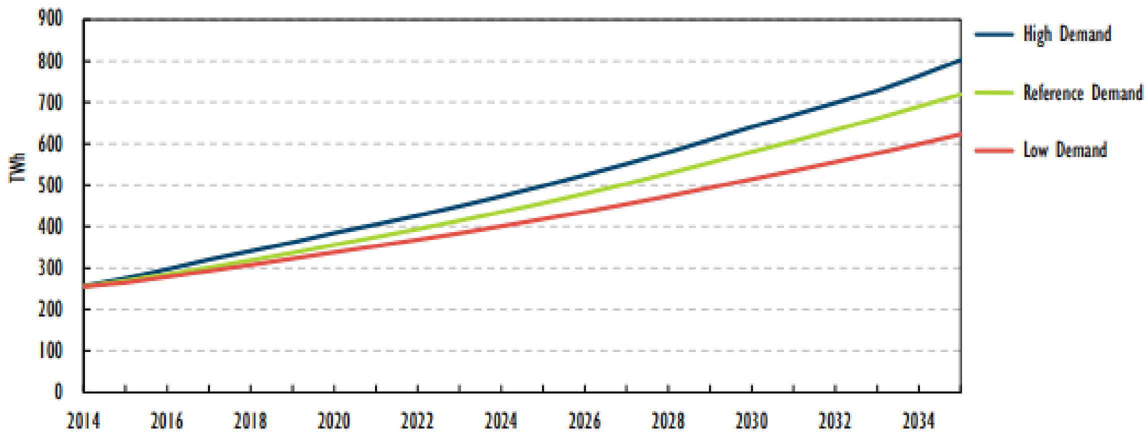


Figure 3. Electricity demand projection, 2014-2035 (MENR, 2014)

At the beginning of each day electricity distribution companies forecast their hourly demand, and notify the regulation authority about their demand forecasts for the next day. These forecasts are essential inputs for electricity generation companies about their production level. Energy Exchange Istanbul – EXIST publishes electricity load forecast data on its website (https://seffaflik.epias.com.tr/transparency/tuketim/tahmin/yuk-tahmin-plani.xhtml). The data show total hourly demand for the next day on an hourly basis. Report dates start from the beginning of 2012.

Electricity demand forecasting is the basis for energy investments, and it plays a vital role in developing countries such as in Turkey. Overestimated demand would lead to idle capacity, which in turn leads to financial resource waste. On the other hand, underestimation would cause energy outages. Therefore, forecasting electricity demand with good accuracy is really important in order to avoid negative consequences.

The remainder of the project is organized as follows: Section 2 summarizes the related literature. We provide results of exploratory data analysis that we performed on data sets in Section 3. Then, in Section 4, machine learning algorithms are applied for forecasting. Lastly, comperative results of machine learning algorithms are discussed in Section 5.

2. LITERATURE REVIEW

Numerous research has been done to improve the accuracy of forecasting over the past years. Parametric methods benefit from the mathematical modeling, where the model parameters are estimated using statistical techniques. Parametric demand forecasting methods are classified under two schemes, namely, time series methods and regression methods (Al-Hamadi, 2005). Regression methods are divided into the linear and nonlinear models, where linear regression methods are the most popular ones in forecasting and prediction (Amirnekooei et al., 2012). Similar to time series models, linear regression methods are based on the linearity assumption.

Studies on the estimation of electric energy demand began in the 1960s in Turkey. The State Planning Organization was using simple regression techniques for estimating electric energy at those dates. Similar studies were carried out by the Ministry of Energy and Natural Resources afterwards.

In the literature, there are plenty of important articles about electricity consumption and demand estimation. Most common methods are time series models, regression models, neural networks, ant colony optimization and statistical learning models. Regression and time series analysis are the most popular modeling techniques in electricity consumption forecasting.

Many researchers carry out electricity demand forecasting studies. Gilland (1988) develops an energy demand projection of the world from 2000 to 2020. Ediger and Tatlidil (2002) and Ediger and Camdali (2007) propose approaches that use the analysis of cyclic patterns in historical curves to predict the primary energy demand in Turkey. Yumurtaci and Asmaz (2004) propose an approach based on the population and energy consumption increase rates per capita to calculate the energy demand of Turkey for the period between 1980 and 2050. Sozen et al. (2005) use Artificial Neural Networks (ANN) to predict Turkey's net energy consumption. Toksari (2007) develops an ant colony energy demand estimation model for Turkey. Akay and Atak (2007) propose an approach using grey prediction with rolling mechanism to predict the Turkey's total and industrial electricity consumption. Murat and Ceylan (2006) state that modeling the energy consumption may be carried out with ANNs. Sozen and Arcaklioglu (2007) develop the energy sources estimation equations in order to estimate the future projections and make correct investments in Turkey using an ANN approach. Hamzacebi (2009) uses ANNs with time series structure to predict Turkish electricity consumption.

2.1. Forecasting Models

Selection of the model to use in forecasting the electricity demand is an important issue. Depending on the nature of the factors affecting the forecast, one method may outperform the others. In this context, comparative results of the models used for forecasting energy demand of Turkey and the comparative evaluation of the Simple Exponential Smoothing, Holt's Linear Trend and Auto Regressive Integrated Moving Average models are examined. These studies contain historical demand data to build a forecasting model and predict future demand.

2.1.1. Simple Exponential Smoothing

Exponential smoothing is an intuitive forecasting method that weights the observed time series unequally. Recent observations are weighted more heavily than remote observations. The unequal weighting is accomplished by using one or more smoothing parameters, which determine how much weight is given to each observation (Li et al. 2008). The simplest technique of this type, simple exponential smoothing (SES), is appropriate for a series that moves randomly above and below a constant mean (stationary series). It has no trend and no seasonal patterns (Yorucu, 2003).

Exponential smoothing model is a popular method in time series analysis. This popularity comes from its simplicity, computational efficiency, responsiveness to changes in the process of forecasting and sufficient accuracy. In general, exponential smoothing is an inexpensive technique that results high accuracy in a wide variety of implementations. Also, data-storage and computing requirements are minimum, which makes exponential smoothing suitable for many applications.

The formula for simple exponential smoothing is;

$$S_t = \alpha *X_t + (1- \alpha)*S_{t-1}$$

When applied recursively to each successive observation in the series, each new smoothed value (forecast) is computed as the weighted average of the current observation and the previous smoothed observation; the previous smoothed observation was computed in turn from the previous observed value and the smoothed value before the previous observation, and so on. Thus, in effect, each smoothed value is the weighted average of the previous observations, where the weights decrease exponentially depending on the value of parameter (α). If it is equal to 1 (one) then the previous observations are ignored entirely; if it is equal to 0 (zero), then the current observation is ignored entirely, and the smoothed value consists entirely of the previous smoothed value (which in turn is computed from the smoothed observation before it, and so on; thus all smoothed values will be equal to the initial smoothed value S_0). In-between values will produce intermediate results (Kalekar, 2004).

2.1.2. Holt’s Linear Trend Method

Holt’s linear method (also knows as Double exponential Smoothing) is an extension to the Simple Exponential Smoothing algorithm originally designed for time series without trend and seasonal patterns. It is used when there is a trend in dataset. Exponential smoothing with a trend works similarly with simple smoothing, except that two components must be updated each period - level and trend. The level is a smoothed estimate of the value of the data at the end of each period. The trend is a smoothed estimate of average growth at the end of each period. Holt’s model is good for non-seasonal data with a trend.

Holt's Linear Trend computes an evolving trend equation through the data using a special weighting function that places the greatest emphasis on the most recent time periods. Instead of the global trend equation of the least squares trend algorithm, this technique uses a local trend equation. The forecasting equation changes from period to period.

The forecasting algorithm uses following formulas:

$$\alpha_t = \alpha Xt + (1 - \alpha)(\alpha_{t-1} + b_{t-1})$$
$$b_t = \beta(\alpha_t - \alpha_{t-1}) + (1 - \beta) b_{t-1}$$

Here α and β are smoothing constants, which are between zero and one. Again, a_t gives the y-intercept (or level) at time t, while b_t is the slope at time t.

The forecast at time T for the value at time T+k is $a_T + b_T k$.

2.1.3. Auto Regressive Integrated Moving Average - ARIMA

One of the most important and widely used time series models is the ARIMA (auto regressive integrated moving average) model. The popularity of the ARIMA model is due to its use of statistical properties as well as the well-known Box & Jenkins methodology and GM (Grey model) in the modeling process (Zhang 2003). There are other models used in forecasting energy consumption in economies such as multiple regression models and artificial neural network models. There are four steps involved in these models, and these steps have been explained by Ajith and Baikunth (2001) as follows:

1. *Model identification*: Attain stationary, and temporarily identify patterns and model components by using graphs, statistics, autocorrelation function, partial autocorrelation functions, transformations, etc.
2. *Parameter estimation*: Find the model coefficients by applying least squares, maximum likelihood methods or other techniques.
3. *Model diagnostics*: Check if the model is valid. If it is, then accept the model; otherwise go back to step 1.
4. *Forecast verification and reasonableness*. Several statistical techniques and confidence intervals determine the validity of forecasts and track model performance to detect out of control situations."

ARIMA is originated from the autoregressive model and the moving average model. ARIMA model can be used when the time series is stationary, and there is no missing data in the time series.

3. EXPLORATORY DATA ANALYSIS

Energy Exchange Istanbul – EXIST, publishes load forecast regarding total hourly demand for the next physical day on its website (https://seffaflik.epias.com.tr/transparency/tuketim/tahmin/yuk-tahmin-plani.xhtml). This data has a vital importance to predict supply and demand equilibrium. The accuracy of forecasting has a great impact on operational and managerial decisions. Electricity load forecasting is an important process that can increase both the revenues and efficiency of the electricity generation and distribution companies.

The dataset is extracted from EXIST and it contains hourly load volumes between June 2016 and May-2018. Figure 4 provides the descriptive information of the dataset.

```
>>> df.describe()
               Load
count  17520.000000
mean   32684.527397
std     4923.295087
min    17800.000000
25%    28700.000000
50%    33000.000000
75%    36100.000000
max    46500.000000
```

```
>>> print(df.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17520 entries, 0 to 17519
Data columns (total 2 columns):
Date    17520 non-null datetime64[ns]
Load    17520 non-null int64
dtypes: datetime64[ns](1), int64(1)
memory usage: 273.8 KB
None
```

Figure 4. Descriptive Information of Dataset

There are 17,520 instances in this dataset and each row represents a load volume of related hours. Minimum load volume is 17,800 MWh and the maximum load volume is 46,500 MWh along with average volume of 32,685 MWh. There is no missing value on the dataset but values were “0” for three days. These values are replaced with the mean of total actual load volumes, 32,685MWh, in data cleaning process. Afterwards, “Hour” attribute is extracted from the dataset since it will be redundant in our analysis. Figure 5 gives our datasets’ head and tail after data pre-processing.

```
>>> df.head(6)
        Date  Load Forecast (MWh)
0 2016-06-01                27800
1 2016-06-01                26600
2 2016-06-01                25600
3 2016-06-01                25200
4 2016-06-01                24900
5 2016-06-01                24400
```

```
>>> df.tail(6)
            Date  Load Forecast (MWh)
17514 2018-05-31                34800
17515 2018-05-31                34700
17516 2018-05-31                34300
17517 2018-05-31                34900
17518 2018-05-31                34500
17519 2018-05-31                33200
```

Figure 5. Glimpse of Dataset

In order to see overall trend in our dataset, we have to aggregate our dataset on a daily level. We can take mean values of each day and plot two years data in a single graph, as in Figure 6.

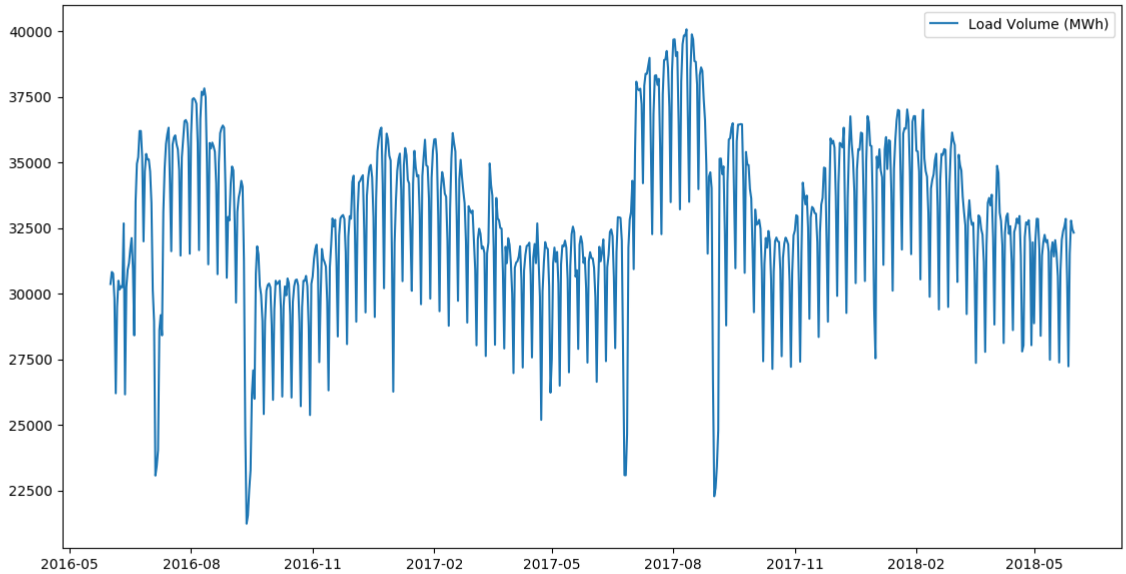


Figure 6. Trend over the period between June-2016 and May-2018

According to Figure 6, load volumes have seasonal trends. Volumes peak during summer time and decrease sharply in autumn seasons. Then, there is a slight increase until winter period, and volumes decrease slowly throughout to spring season. The reason of high load volumes in summer times is heavily usage of air conditioners and agricultural irrigations in Turkey. On the other hand, load volumes are quite high during winter time because of increases in heating demand.

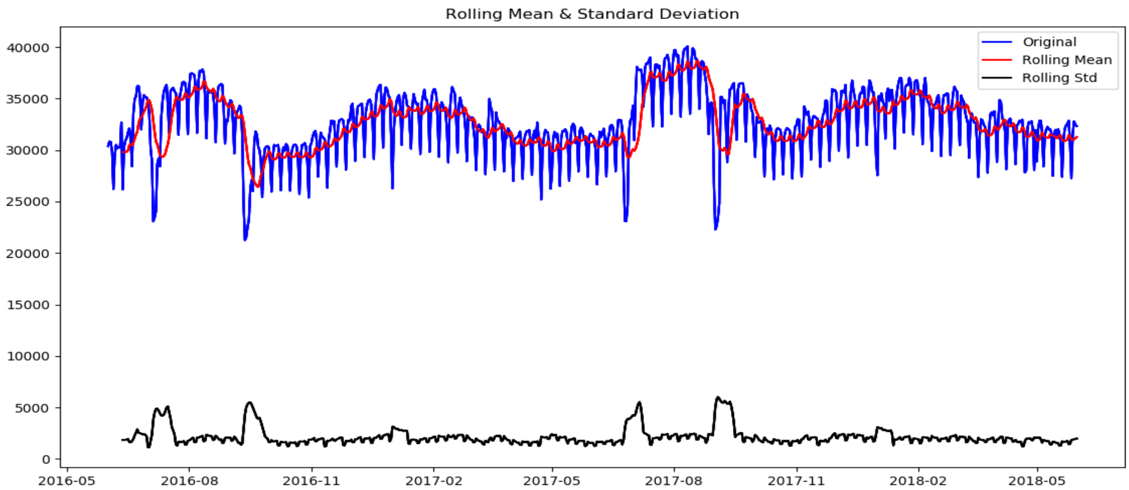


Figure 7. Rolling mean and standard deviation of dataset

In Figure 7 we notice that time series is not stationary. The major reason seems to be seasonality since there are variations at specific time frames. Stationarity is the most common assumption in time-series models. To overcome this assumption, some models remove seasonality effects from series, apply forecasting models, and then convert forecasted values into original scale. On the other hand, some models take account of seasonalities in time series models.

In order to train our model and test our results, the dataset is split in two sets; namely train and test datasets. The training set contains known outputs; models learn from this data and make generalizations for new data. With a test set, we can evaluate our model's prediction. We have split our dataset as given in Figure 8.

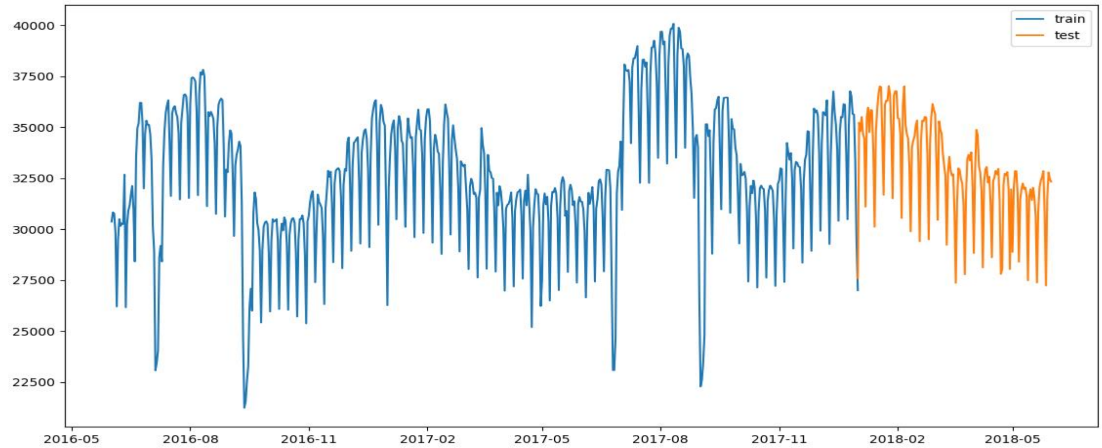


Figure 8. Train&Test split of dataset

4. MACHINE LEARNING METHODS

4.1. Simple Exponential Smoothing

Naive forecasting places 100% weight on the most recent observation and moving averages place equal weight on k values. Exponential smoothing allows for weighted averages where greater weight can be placed on recent observations and less weight on older observations. Exponential smoothing methods are intuitive, computationally efficient, and generally applicable to a wide range of time series. Forecasts are calculated using weighted averages where the weights decrease exponentially as observations come from further in the past, the smallest weights are associated with the oldest observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots$$

where $0 \leq \alpha \leq 1$ is the smoothing parameter. The one-step-ahead forecast for time T+1 is a weighted average of all observations in the series $y_1, \ldots, y_T$. The rate at which the weights decrease is controlled by the parameter α. As we can see, $1-\alpha$ is multiplied by the previous expected value $\hat{y}_{x-1}$, which makes the expression recursive. This is why this method is called exponential. The forecast at time t+1 is equal to a weighted average between the most recent observation y_t and the most recent forecast $\hat{y}_{t|t-1}$.

In Figure 9, we can see Simple Exponential Smoothing (SES) Model's outcome. Predicted values can be seen in green line – SES Model, they are close to mean values of recent observations in train dataset.

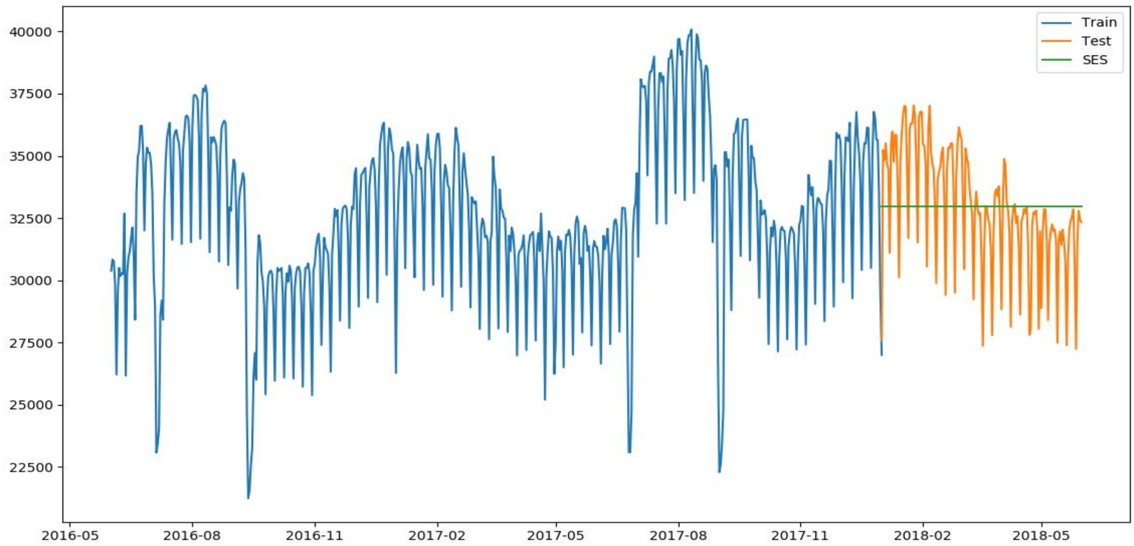


Figure 9. Simple Exponential Smoothing Model

We will use Root Mean Square Error (RMSE) metric in order to evaluate our predictions. RMSE is the standard deviation of the residuals (prediction errors). Residuals are a measure of distance between regression line and data points. In other words, it compares a predicted value and observed or known value. The RMSE of our test dataset is 2389 with SES Model. RMSE has the same unit as load values in our dataset. Our aim is to minimize RMSE scores.

Another metric for evaluating our method is R Squared, it provides an indication of the goodness of fit of a set of predictions to the actual values. In statistical literature, this measure is called the coefficient of determination. Score ranges between 0 and 1 (no-fit and perfect fit). R Squared score is -0.0003 in Simple Exponential Smoothing (SES) Model. If we have used the mean value instead of using SES Model for our prediction, we would get better R Squared score, i.e 0. We will try another methods to decrease RMSE and increase R Squared score.

4.2. Holt's Linear Trend Method

Trend is the general pattern of values that we montitor over a time period. Each time series dataset can be decomposed into it's components, which are Trend, Seasonality and Residual. Datasets with a trend can use Holt's linear trend method for forecasting. In this case we can see that there is a trend in some periods over the year as given in Figure 10.

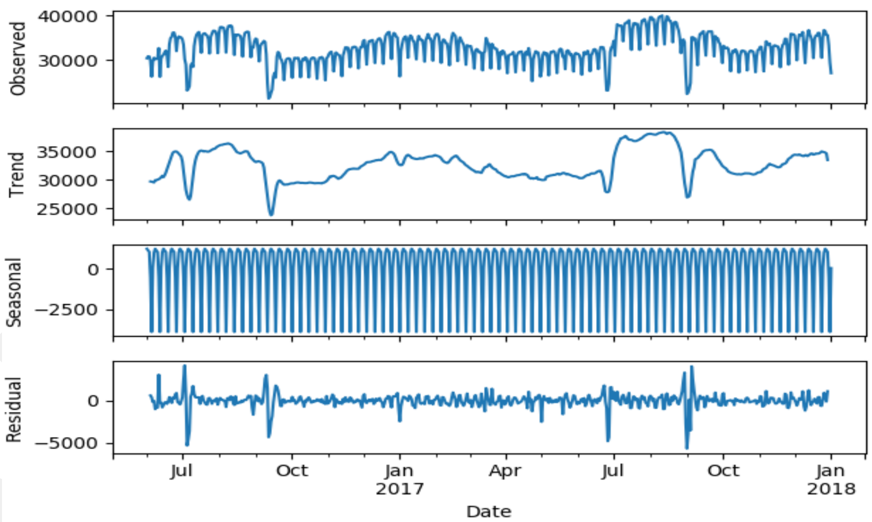


Figure 10. Trend analysis of train dataset

Holt extends simple exponential smoothing to enable forecasting of data with a trend. It is a exponential smoothing method applied to both level (the average value in the series) and trend. To express this in mathematical notation we now need three equations: level, trend and forecast in order to combine the level and trend, and get expected results. As with simple exponential smoothing, the level equation here indicates that it is a weighted average of observation and the within-sample one-step-ahead forecast. Level and trend equations are added in order to generate the forecast equation:

Level equation $Lt = \alpha Yt + (1 - \alpha)(Lt - 1 + bt - 1)$

Trend equation $bt = \beta(Lt - Lt - 1) + (1 - \beta)bt - 1$

Forecast equation $Ft + m = Lt + btm$

In Figure 11, Holt's Linear Trend Model's outcome can be seen. Predicted values, as shown in green line – Holt_linear, decrasing slightly due to the trend in the dataset.

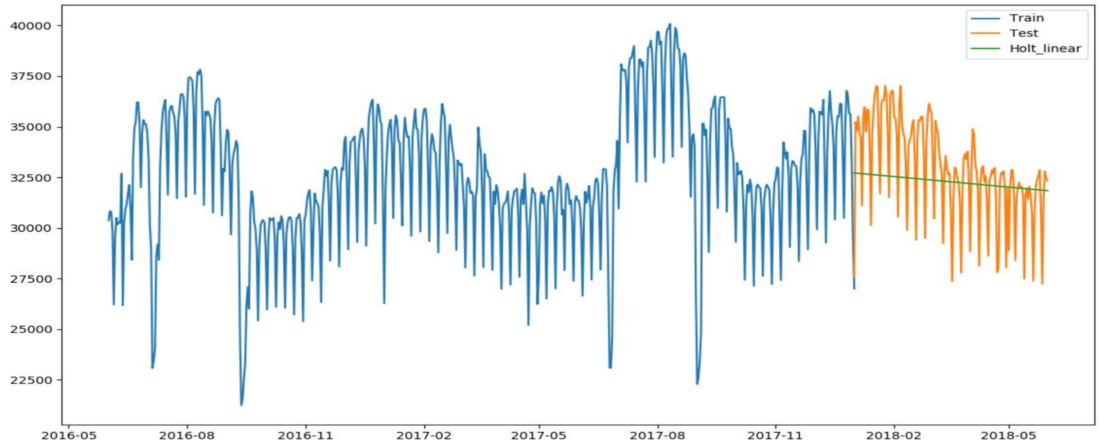


Figure 11. Holt's Linear Trend Model

The RMSE score of our prediction is 2273 with Holt's Linear Trend Method. We can see that we received better prediction results compared to Simple Exponential Smoothing Method, because trend function is added in our forecast equation. R Squared score is 0.09 with this model, which is better than SES Model as well.

4.3. Auto Regressive Integrated Moving Average - ARIMA

Auto Regressive Integrated Moving Average – ARIMA is a very popular time series model in data science. Exponential smoothing models are based on trend and seasonality in the data. However, ARIMA model can demonstrate the correlations in the data with each other. Seasonal ARIMA (SARIMA) is an improvement method over ARIMA. Seasonality in the time series is a regular pattern of changes that recur over S time periods, where S is the number of time periods until the pattern recurs again. Our dataset has monthly figures, so in this case S=12 (no of months in a year), and mean load volume is not constant and it varies by month. Seasonality usually leads the series to be nonstationary because the average values at some specific times within the seasonal span (e.g. months, weeks) may be different than the average values at other times. When there is seasonality in a dataset, mean of the observations is not constant and it changes according to a cyclical pattern as shown in Figure 12.

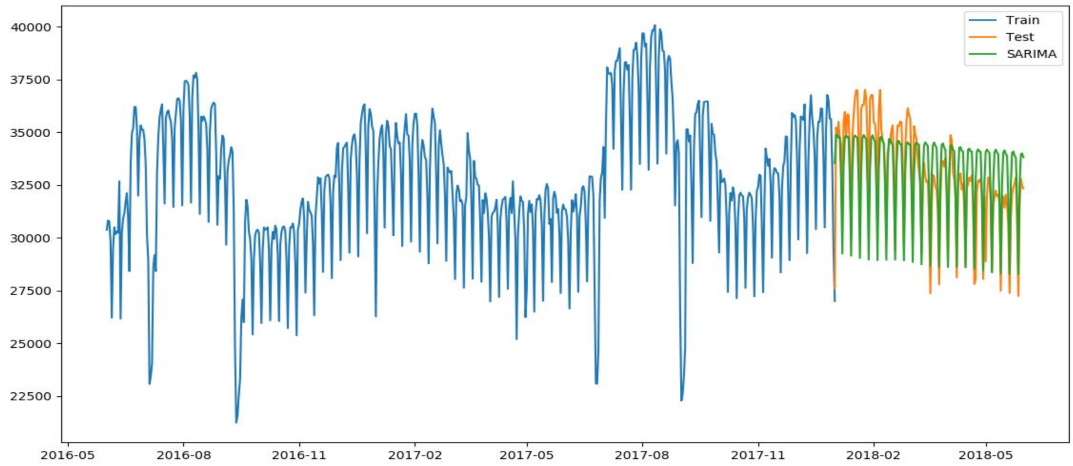


Figure 12. SARIMA Model

In Figure 12, we can see that actual load volumes are decreasing slightly between January and May (prediction period) in 2018. Our model's prediction is also in line with this trend. When we take a look at prediction period in 2017 part, we can notice major similarities between 2017 and 2018 volumes.

RMSE score is 1565 with SARIMA Model. Using trend and seasonality function gives better prediction results in our dataset. R Squared score is 0.57 and this is highest score among our models.

5. CONCLUSION

In this project, hourly electricity load volumes are used in order to predict future values. The data set consist of hourly load volumes between 01.06.2016 and 31.05.2018. The importance of demand forecasting is explained in Section 1 and related literature is summarized in Section 2.

We started with exploratory data analysis in order to explore the characteristics of our dataset. Descriptive information is given in Section 3. In data cleaning process, null values are replaced with mean value and some attributes are extracted since they were redundant. Hourly load volumes are aggregated to daily level, so we can see the trend and seasonality functions more clearly. The dataset is splitted into 2 sets, train and test, in order to train the models and test our results.

In Section 4, machine learning methods are applied to dataset with Python programming language. First method is called Simple Exponential Smoothing - SES. In this method, recent observations are weighted greater than older observations. The SES method doesn't account for trend and seasonality.

Second method is called Holt's Linear Trend. This method is an extension to SES method and take acount of trend function. Since there is trend pattern in our data set, Holt's Linear Trend method gives better prediction results.

Third method is called Auto Regressive Integrated Moving Average - ARIMA. ARIMA model is originated from the autoregressive and moving average models. Seasonal ARIMA method is used in our project in order to take account both trend and seasonality.

METHOD	RMSE	R Squared	Trend	Seasonality
Simple Exponential Smoothing	2389	-0.0003	No	No
Holt's Linear Trend Method	2273	0.09	Yes	No
Seasonal Auto Regressive Integrated Moving Average	1565	0.57	Yes	Yes

Table 1. Evaluation Metrics

In Table 1, our test results are given. The models are starting from simple to more complex. When we check the accuracy scores, SARIMA gives best results with lowest RMSE and highest R Squared score among our methods. The reason behind this fact is, SARIMA model take account of both trend and seasonality.

This project shows us, electricity demand can be modeled using machine learning algorithms, and the models can be used to predict future electiricty demand. Models that take into account of trend and seasonality functions for electricity demand forecast would give better accuracy scores in future studies.

APPENDIX

PYTHON SCRIPT

```
#Libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import xlrd
from statsmodels.tsa.stattools import adfuller
from math import sqrt
from sklearn.metrics import mean_squared_error
from statsmodels.tsa.api import SimpleExpSmoothing, Holt
import statsmodels.api as sm
from sklearn.metrics import r2_score

#EDA

df = pd.read_excel('train.xlsx')
df.head(6)
df.tail(6)
df.describe()
print(df.info())

df.Timestamp = pd.to_datetime(df.Date,format='%d-%m-%y')
df.index = df.Timestamp
df = df.resample('D').mean()
plt.figure(figsize=(10, 8))
plt.plot(df, label = 'Load Volume (MWh)')
plt.legend()

def test_stationarity(timeseries):
    rolmean = pd.rolling_mean(timeseries, window=12)
    rolstd = pd.rolling_std(timeseries, window=12)
    orig = plt.plot(timeseries, color='blue', label='Original')
    mean = plt.plot(rolmean, color='red', label='Rolling Mean')
    std = plt.plot(rolstd, color='black', label='Rolling Std')
    plt.legend(loc='best')
    plt.title('Rolling Mean & Standard Deviation')
    plt.show(block=False)

test_stationarity(df)
```

```
#train&test split

train=df[0:13898]
test=df[13898:]

train.Timestamp = pd.to_datetime(train.Date,format='%d-%m-%y')
train.index = train.Timestamp
train = train.resample('D').mean()
test.Timestamp = pd.to_datetime(test.Date,format='%d-%m-%y')
test.index = test.Timestamp
test = test.resample('D').mean()
plt.figure(figsize=(10, 8))
plt.plot(train, label = 'train')
plt.plot(test, label = 'test')
plt.legend()

#Simple Exponential Smoothing

y_hat_avg = test.copy()
fit2 = SimpleExpSmoothing(np.asarray(train)).fit(smoothing_level=0.15,optimized=False)
y_hat_avg['SES'] = fit2.forecast(len(test))
plt.figure(figsize=(16,8))
plt.plot(train, label='Train')
plt.plot(test, label='Test')
plt.plot(y_hat_avg['SES'], label='SES')
plt.legend(loc='best')
plt.show()

rms_test = sqrt(mean_squared_error(test, y_hat_avg.SES))
print(rms_test)
r2_score(test,y_hat_avg['SES'])

#Holts linear

sm.tsa.seasonal_decompose(train).plot()
result = sm.tsa.stattools.adfuller(train)
plt.show()

y_hat_avg = test.copy()
fit1 = Holt(np.asarray(train)).fit(smoothing_level = 0.16,smoothing_slope = 0.02)
y_hat_avg['Holt_linear'] = fit1.forecast(len(test))
plt.figure(figsize=(16,8))
plt.plot(train, label='Train')
plt.plot(test, label='Test')
plt.plot(y_hat_avg['Holt_linear'], label='Holt_linear')
plt.legend(loc='best')
plt.show()
```

```
rms_test = sqrt(mean_squared_error(test, y_hat_avg.Holt_linear))
print(rms_test)
r2_score(test,y_hat_avg['Holt_linear'])

#ARIMA

y_hat_avg = test.copy()
fit1 = sm.tsa.statespace.SARIMAX(train, order=(2, 1, 4),seasonal_order=(0,1,1,7)).fit()
y_hat_avg['SARIMA'] = fit1.predict(start="2018-01-01", end="2018-05-31",
dynamic=True)
plt.figure(figsize=(16,8))
plt.plot( train, label='Train')
plt.plot(test, label='Test')
plt.plot(y_hat_avg['SARIMA'], label='SARIMA')
plt.legend(loc='best')
plt.show()

rms_test = sqrt(mean_squared_error(test, y_hat_avg.SARIMA))
print(rms_test)
r2_score(test,y_hat_avg['SARIMA'])
```

TABLES AND GRAPHS

Graph List

Table List

REFERENCES

- The Oxford Institute for Energy Studies, Gas Supply Changes in Turkey, 2018
- Energy Policies of IEA Countries, Turkey, 2016 Review
- Electricity generation by source, 1973-2015 (IEA (2016), Energy Balances of OECD Countries 2016, www.iea.org/statistics/)
- Electricity consumption by sector, 1973-2014 (IEA (2016a), Electricity Information 2016, www.iea.org/statistics/)
- Electricity demand projection, 2014-2035 (MENR (2014)
- Al-Hamadi HM. SAS. Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. Electr Power Syst Res 2005;74: 353e61
- Amirnekooei K, Ardehali MM, Sadri A. Integrated resource planning for Iran: development of reference energy system, forecast, and long-term energyenvironment plan. Energy 2012;46(1):374e85
- Gilland B. Population, economic growth, and energy demand, 1985–2020. Popul Dev Rev 1988;14(2):233–44
- Ediger VS, Tatlidil H. Forecasting the primary energy demand in Turkey and analysis of cyclic patterns. Energy Convers Manage 2002;43:473–87
- Ediger VS, Camdali U. Camdali, energy and exergy efficiencies in Turkish transportation sector, 1988–2004. Energy Policy 2007;35:1238–44
- Yumurtaci Z, Asmaz E. Electric energy demand of Turkey for the year 2050. Energy Source 2004;26:1157–64
- Sozen A, Arcaklioglu E, Ozkaymak M. Modelling of the Turkey's net energy consumption using artificial neural network. Int J Comp Appl Technol 2005;22(2/3)
- Toksari MD. Ant colony optimization approach to estimate energy demand in Turkey. Energy Policy 2007;35:3984–90
- Akay D, Atak M. Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. Energy 2007;32(9):1670–5

- Murat YS, Ceylan HH. Use of artificial neural networks for transport energy demand modeling. Energy Policy 2006;34:3165–72

- Sozen A, Arcaklioglu E. Prospects for future projections of the basic energy sources in Turkey. Energy Source Part B: Econ Plan Policy 2007;2(2):183–201

- Hamzacebi C. Forecasting of Turkey's net electricity energy consumption on sectoral bases. Energy Policy 2007;35:2009–16

- LI, Z. P. – YU, H. – LIU, Y. C. – LIU, F. Q.: An Improved Adaptive Exponential Smoothing Model for Short Term Travel Time Forecasting of Urban Arterial Street, Acta automatica sinica, Vol. 34, No. 11, 1404–1409, 2008.

- YORUCU, V.: The Analysis of Forecasting Performance by Using Time Series Data for Two Mediterranean Island, Review of Social, Economic & Business Studies, Vol. 2, 175–196, 2003

- Prajakta S. Kalekar, Time series Forecasting using Holt-Winters Exponential Smoothing, 2004

- Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003;50:159e75)

- Ajith, A., Baikunth, N. (2001), A neuro-fuzzy approach for modelling electricity demand in Victoria. Applied Soft Computing, 1(2), 127-138