



# Forecasting with Ensemble Methods: An Application Using Fashion Retail Sales Data


Orkun Berk Yüzbaşıoğlu

# Summary

- In this project, ensemble machine learning models are used to predict short term store sales of a fashion retailer.
- Sales forecasts of various products at different stores are generated for a span of three months with bagging tree regressor, random forest regressor, and gradient boosting regressor algorithm.
- Algorithms are trained and evaluated with real past sales data of a Turkish fashion retailer.

# 1. Literature Review on Forecasting Applied for Fashion Industry and Fashion Retail Sales (Theoretical & Academic)



 Makaleler

Yaklaşık 16.900 sonuç bulundu (0,11 sn)

Tüm zamanlar

2020 yılından beri

2019 yılından beri

2016 yılından beri

Özel aralık...

—

Alakaya göre sırala

Tarihe göre sırala

Herhangi bir dil

Türkçe sayfalarda ara

☒ patentleri içer

☒ alıntıları

☒ Uyarı oluştur

Exploring the use of deep neural networks for sales **forecasting** in **fashion retail**

ALD Loureiro, [VL Miguéis](#), LFM da Silva - *Decision Support Systems*, 2018 - Elsevier

... with a stock-out situation, they might decide to shop in a different **retailer** [6]. Additionally ... by evaluating the efficiency of DNN to perform sales predictions in the **fashion retail** industry when ... This research work has the objective of predicting the sales of new **fashion** products using ...

☆ 99 Alıntılanma sayısı: 35 İlgili makaleler 3 sürümün hepsi

[PDF] e-tarjome.com

Retail **forecasting**: Research and practice

R Fildes, S Ma, [S Kolassa](#) - *International Journal of Forecasting*, 2019 - Elsevier

☆ 99 Alıntılanma sayısı: 21 İlgili makaleler 12 sürümün hepsi

[PDF] lancs.ac.uk

A data-driven **forecasting** approach for newly launched seasonal products by leveraging **machine-learning** approaches

M Kharfan, VWK Chan, [TF Efendigil](#) - *Annals of Operations Research*, 2020 - Springer

... the purchasing power of customers that make the **forecasting** process extremely complicated in the **fashion retail** industry ... Throughout the **fashion** supply chains including many entities such as raw materials suppliers, manufacturers, distributors and **retailers**, the orders being ...

☆ 99 4 sürümün hepsi

[PDF] icm.edu.pl

Intelligent **retail forecasting** system for new **clothing** products considering stock-out


H Huang, Q Liu - *Fibres & Textiles in Eastern Europe*, 2017 - yadda.icm.edu.pl

... Given the data of new **clothing** products is limited, a data selection process is integrated in this ... In addition, the life cycle of fast-**fashion** products is short, usually 6-10 weeks [16] ... The main reason might be that **retail** stores have shorter business hours on Sunday in North America ...

☆ 99 Alıntılanma sayısı: 9 İlgili makaleler 4 sürümün hepsi

[PDF] icm.edu.pl

# 1. Literature Review on Forecasting Applied for Fashion Industry and Fashion Retail Sales (Theoretical & Academic)

**MEF**  
Kütüph@ne  
bilgiye sınırsız erişim

Kütüphaneciye Sor

[f](#) [t](#) [y](#) [i](#) [p](#) [m](#)

EN


AraştırmaHakkımızdaHizmetlerimizBültenFormlarKim Kimdir

VeritabanlarıKütüphane Hesabım




















## Abone Veritabanları

[Abone Veritabanları](#) [Açık Erişim Veritabanları](#) [Deneme Veritabanları](#)

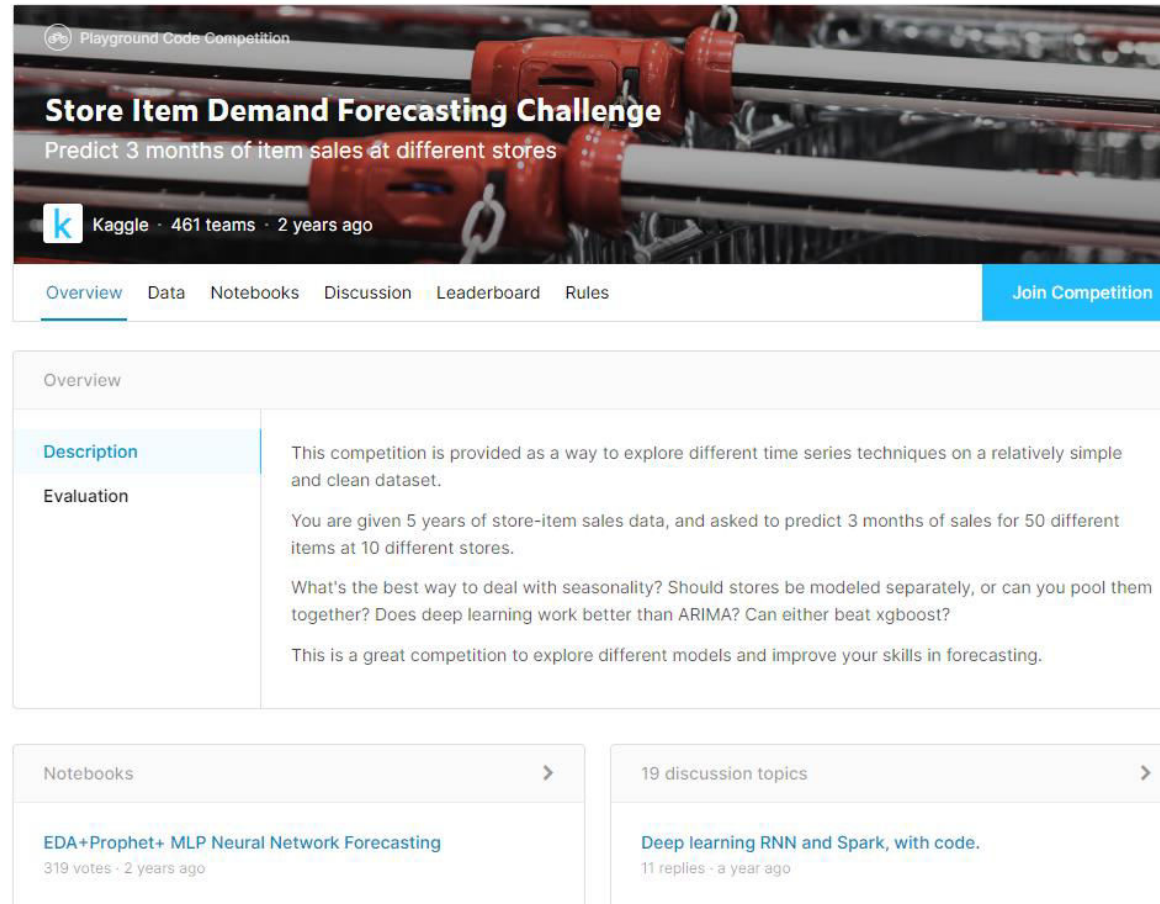
### Abone Veritabanları

MEF Üniversitesi öğrencileri, akademik ve idari personeli yararlanabilmektedir. 

[Kampus dışından veritabanlarına erişmek için tıklayınız](#) [Off - Campus Logout](#)

<a href="#">A to Z Dergi Listesi (EBSCO - FDS)</a>	  
 <a href="#">Koronavirüs (COVID-19) Ücretsiz Araştırma, e-Kitaplar ve Klinik Bilgiler</a> 	  
<a href="#">Academic Focus - China English Academic Journals &amp; Proceedings Database (CEJD)</a>	  
<a href="#">Beck-online</a>	  
<a href="#">Bloomsbury Design Library</a>	  
<a href="#">Bloomsbury Applied Visual Arts</a>	  

# 1. Literature Review on Forecasting Applied for Fashion Industry and Fashion Retail Sales (Practical)



The screenshot displays the Kaggle competition page for the 'Store Item Demand Forecasting Challenge'. The header features a banner with the competition title and a sub-header 'Predict 3 months of item sales at different stores'. Below the banner, a navigation bar includes links for Overview, Data, Notebooks, Discussion, Leaderboard, and Rules, along with a 'Join Competition' button. The main content area is divided into sections for Description and Evaluation. The Description section explains the competition's purpose and the data provided. The Evaluation section discusses the challenges of forecasting, such as seasonality and model selection. Below the main content, there are two sidebars: one for Notebooks and one for Discussion topics. The Notebooks sidebar lists a notebook titled 'EDA+Prophet+ MLP Neural Network Forecasting' with 319 votes. The Discussion topics sidebar lists a topic titled 'Deep learning RNN and Spark, with code.' with 11 replies.

Playground Code Competition

## Store Item Demand Forecasting Challenge

Predict 3 months of item sales at different stores

Kaggle · 461 teams · 2 years ago

[Overview](#) [Data](#) [Notebooks](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Join Competition](#)

### Overview

Description	Evaluation
This competition is provided as a way to explore different time series techniques on a relatively simple and clean dataset.	You are given 5 years of store-item sales data, and asked to predict 3 months of sales for 50 different items at 10 different stores.
	What's the best way to deal with seasonality? Should stores be modeled separately, or can you pool them together? Does deep learning work better than ARIMA? Can either beat xgboost?
	This is a great competition to explore different models and improve your skills in forecasting.

### Notebooks

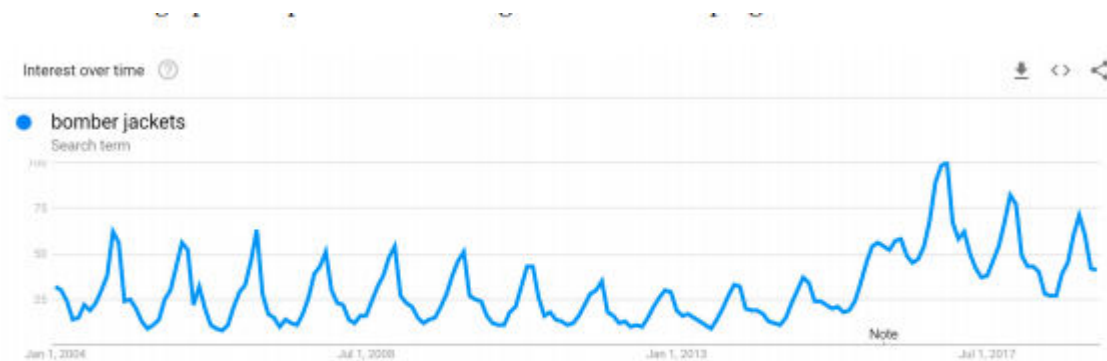
EDA+Prophet+ MLP Neural Network Forecasting  
319 votes · 2 years ago

### 19 discussion topics

Deep learning RNN and Spark, with code.  
11 replies · a year ago

## 2. Data Retrieval

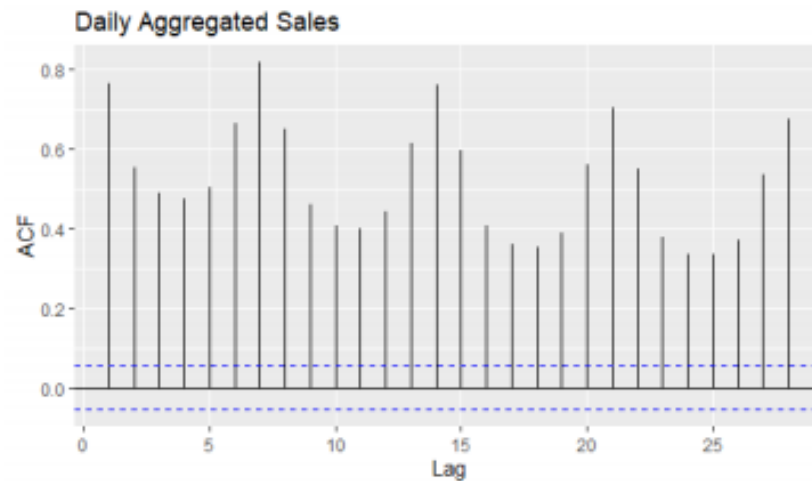
- Sales Data (From RDBMS)
- Past Weather (NOAA)
- Weather Forecast (scrapy or from API)
- Special Dates (calendar package)
- Google Trends (pytrends package)



**Figure 3.** Google Trends for bomber jackets (January 2004–February 2019) (Data Source: Google Trends, 1 February 2019).

# 3. EDA

- To look for patterns in data
- For feature engineering
- For time series data: ACF, Line Charts



**Figure 2. ACF plot of Daily Sales**

Figure. ACF Plot

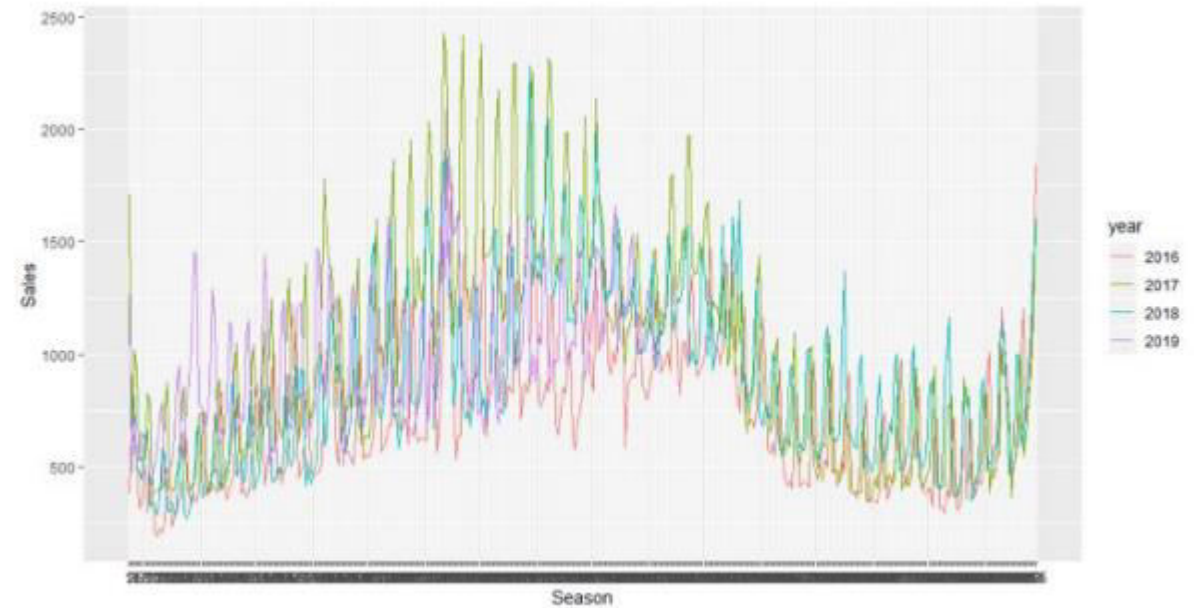


Figure. Line Chart

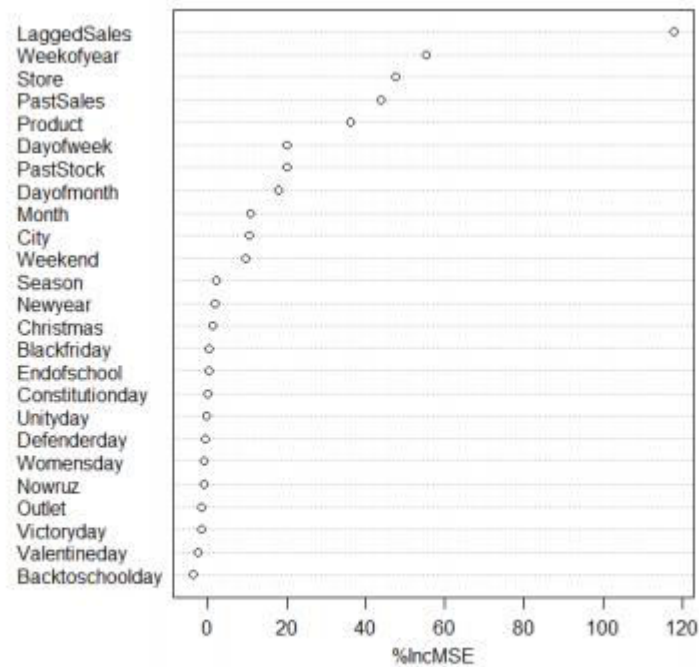
## 4. Feature Engineering

- Lagged Sales Features (from ACF plot sales of the same day from the week before last)
- Special Dates (such as Valentine's day, Republic Day (29 October))
- Geographic Features (city, country)
- Date Time Features:
  - Day of Week (1-7)
  - Day of Month (1-31)
  - Week of Year (1-52)\*
  - Month (1-12),
  - Weekend (0, 1)



# 5. Feature Selection

- 1. Feature Importance from Random Forest



**Figure 8. Variable Importance from Bagging Regression Tree**

## 5. Feature Selection

- Scikit-Learn's **Feature Selection** Class ([https://scikit-learn.org/stable/modules/feature\\_selection.html](https://scikit-learn.org/stable/modules/feature_selection.html))
- **Removing Features with low variance:** *[VarianceThreshold](#) is a simple baseline approach to feature selection. It removes all features whose variance doesn't meet some threshold. By default, it removes all zero-variance features*
- **Recursive Feature Elimination with an estimator:** *Given an external estimator that assigns weights to features (e.g., the coefficients of a linear model), recursive feature elimination ([RFE](#)) is to select features by recursively considering smaller and smaller sets of features.*

## 6. Model Training

- Train-Validation-Test Split: No hyper-parameter tuning based on test Set
- Hyper-parameter's of model: From package documentation (scikit-learn, caret, lightgbm, xgboost, catboost, ngboost, tensorflow)
- Hyper-parameter tuning:
  - RandomizedSearch
  - GridSearch

# 7. Evaluating Results

- Metrics to compare different Models & Settings:  
RMSE, MAPE, MSE
- Visualizing Results:
  - Scatter Plots (y-test vs y-pred)
  - Residual/Error Plots (error vs y-test)

	Test Set $R^2$	RMSE
Bagging Regression Tree	70.8 %	8.05
Random Forest Regressor	72.88 %	7.71
Gradient Boosted Regressor	71.07%	8.01
Linear Regression	70.03%	8.06

