FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Predicting demand in fashion retailing: a data analytics approach

Tiago Manuel Martins Grilo

MASTER'S THESIS

Mestrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Vera Lucia Miguéis Oliveira e Silva

February 25, 2022

Resumo

No contexto da venda a retalho, uma determinação errónea das quantidades a comprar de cada artigo dos fornecedores, seja por excesso ou por defeito, pode resultar em custos desnecessários de armazenamento ou vendas perdidas, respetivamente. Ambas as situações têm de ser evitadas por empresas, o que promove a necessidade de determinar as quantidades de compra eficientemente.

Presentemente, empresas recolhem grandes quantidades de dados relativos às suas vendas e características dos seus produtos. No passado, essa informação era raramente analisada e integrada no processo de tomada de decisão. Contudo, o aumento da capacidade de processamento de informação promoveu o uso de análise de dados como meio de obter conhecimento e apoiar tomadores de decisão no alcance melhor resultados.

Este projeto tem a finalidade de explorar o uso de técnicas de data mining para otimizar as quantidades a comprar de cada produto vendido por uma empresa de venda a retalho de moda, resultando no desenvolvimento de um modelo que utiliza dados históricos de vendas de produtos com características semelhantes para prever a quantidade que a empresa potencialmente venderá de novos produtos. O projeto utilizará como caso de estudo um empresa portuguesa de venda a retalho de moda.

Keywords: Previsão da procura, Fashion retail

Abstract

In the retail context, an erroneous determination of the amounts to buy of each article from the suppliers, either by excess or defect, can result in unnecessary costs of storage or lost sales, respectively. Both situations should be avoided by companies, which promotes the need to determine purchase quantities efficiently.

Currently companies collect huge amounts of data referring to their sales and products' features. In the past, that information was seldom analyzed and integrated in the decision making process. However, the increase of the information processing capacity has promoted the use of data analytics as a means to obtain knowledge and support decision makers in achieving better business outcomes.

This project aims at exploring the use of data mining techniques to optimize the amounts to buy of each product sold by a fashion retail company, resulting in the development of a model that uses past sales data of the products with similar characteristics to predict the quantity the company will potentially sell from the new products. The project uses as a case study a Portuguese fashion retail company.

Keywords: Demand forecasting, Fashion retail

Agradecimentos

A área académica pode, por vezes, ser muito ingrata na medida em que o conhecimento produzido fica frequentemente retido pela lógica do lucro infinito que não permite o seu pleno usufruto pela sociedade que o produz, por isso o meu muito obrigado à Alexandra Elbakyan pela criação do Sci-Hub, sem o qual nunca teria sido possível elaborar o levantamento do estado da arte que constitui a base desta dissertação.

Uma pessoa com recursos para trabalhar mas sem alguém para a orientar rumo ao caminho certo é como um rio com água mas sem caminho para a levar. Foi para levar as coisas ao rumo certo que contei com a enorme prestabilidade, paciência e amabilidade da minha orientadora, a Professora Vera Miguéis, sem a qual eu nunca teria conseguido ultrapassar a linha de chegada da minha dissertação. O meu enorme obrigado.

Amigo verdadeiro vale mais do que dinheiro e os meus amigos são a confirmação desse provérbio, tanto os meus amigos do coração como os amigos que fui conhecendo por esses cantos cibernéticos. São meus amigos de verdade e estou-lhes grato por tudo.

Uma pessoa sem o calor, o carinho e o apoio da sua família fica desamparada e a minha família foi fundamental para mim neste período tão negro tanto a nível pessoal como a nível global que calhou coincidir em larga medida com a elaboração da presente dissertação. Sem todo o apoio, amor, carinho e paciência deles, eu não seria nada. Um obrigado do tamanho do mundo às pessoas mais importantes da minha vida: o meu pai, a minha mãe, a minha irmã, a minha avó e a minha restante família.

Obrigado,

Tiago Grilo

"All things must pass"

George Harrison

Contents

1 Introduction 1
1.1 Context . . . 1
1.2 Motivation and Objectives . . . 1
1.3 Thesis Outline . . . 2

2 Background 3
2.1 Demand Estimation and Forecasting . . . 3
2.1.1 Models for Demand Estimation and Forecasting . . . 3
2.2 Summary . . . 5

3 Literature Review 7
3.1 Focus . . . 7
3.2 Data . . . 8
3.3 Predictive Variables . . . 10
3.4 Methods Used . . . 11
3.5 Summary . . . 13

4 Data & Methodology 15
4.1 Scope . . . 15
4.2 Available Data . . . 15
4.2.1 Product Data . . . 15
4.2.2 Sales Data . . . 17
4.3 Methodologies . . . 17
4.3.1 Data Preprocessing . . . 17
4.3.2 Data Mining Techniques . . . 18
4.3.3 Model evaluation . . . 22

5 Results & Discussion 25

6 Conclusions 31

References 33

List of Figures

2.1 Architecture of a generic random forest model 5

3.1 Framework of the hybrid intelligent model developed by Wong and Guo 8
3.2 Schematic representation of the methodology followed by Loureiro et al. 12

4.1 Total sales aggregated by product family for the 2015 Spring-Summer season . . 16
4.2 Total sales aggregated by product expectation level for the 2015 Spring-Summer season 17
4.3 Total sales aggregated by store type for the 2015 Spring-Summer season 18
4.4 Total sales aggregated by product fashion level for the 2015 Spring-Summer season 19
4.5 The process of knowledge discovery in databases (KDD) 21
4.6 Fragment of a decision tree generated by the training algorithm 22
4.7 Analogy between biological neural networks and artificial neural networks 23
4.8 Example of a modelled boolean function using a perceptron 24
4.9 Example of feedforward ANN with one hidden layer 24

5.1 Scatter plot of observed sales and predicted sales for the random forest 26
5.2 Scatter plot of observed sales and predicted sales for the deep neural network . . 27
5.3 Scatter plot of observed sales and predicted sales for the support vector machine . 27
5.4 Scatter plot of observed sales and predicted sales for the decision tree 28
5.5 Scatter plot of observed sales and predicted sales for the neural network 28
5.6 Scatter plot of observed sales and predicted sales for the linear regression 29

List of Tables

4.1 Attributes in the company's database 20

5.1 Performance results for the various regression techniques on the testing dataset . 25

Abbreviations

AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CNN	Convolutional Neural Network
DNN	Deep Neural Network
ELM	Extreme Learning Machine
EM	Expectation-Maximization
ETS	Error, Trend and Seasonality
FARIMA	Autoregressive Fractionally Integrated Moving Average
FIS	Fuzzy Inference System
GRU	Gated Recurrent Unit
HS	Harmony Search
KNN	K-Nearest Neighbours
LASSO	Least Absolute Shrinkage and Selection Operator
MAE	Mean Absolute Error
MARS	Multivariate Adaptive Regression Splines
ML	Machine Learning
MSE	Mean Squared Error
RBFNN	Radial Basis Function Neural Network
POS	Point of Sale
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors
SKU	Stock Keeping Unit
SLFN	Single Hidden Layer Feed-Forward Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression

Chapter 1

Introduction

1.1 Context

Due to the high competition in the fashion retail market, a prediction of demand is crucial for any company wanting to be successful. An erroneous determination of the quantity of product to order to suppliers will result in either unnecessary holding costs or lost sales. If the quantity to order is determined in excess, it results in unsold products and storage costs and if the quantity to order is determined in defect, it results in lost sales. This means that, in order to avoid these situations, it is necessary to determine the quantities to be ordered effectively and efficiently.

The sales and demand of a product are highly dependent on multiple factors and variables, making this type of forecasting a rather complex subject. So, accurate demand and sales forecasting is crucial to assist in production planning and business improvement [1].

1.2 Motivation and Objectives

Currently, companies collect large amounts of data related with their products and sales. In the past, these data were seldom analyzed and integrated in the decision-making process. However, with the increase in processing power, a new window of opportunity is opened through the use of data analysis techniques in order to obtain knowledge and support the decision-making process to reach better forecasting results. It is, therefore, important to explore a predictive model that incorporates these data in the demand forecasting process.

Through the usage of data mining techniques, the objective of this project is to develop a predictive model for demand forecasting, infer sales potential and ultimately develop an application that optimizes the quantities of each product of a fashion retailing company to order.

It is in our interest, having a robust forecasting model, to identify the most relevant predictors and features of product demand and sales, because these predictors (or a combination of them) will add value to the model's performance, if chosen wisely.

In order to test the predictive model against real data, a Portuguese fashion retail company will be used as a case study.

1.3 Thesis Outline

This document contains six chapters, including the introduction.

In chapter 2, relevant theoretical background on various subjects and topics explored in this work is provided.

Chapter 3 contains an overview of related works, a literature review and a review of the different types of methods used for demand forecasting.

In chapter 4, the problem and data set associated with the dissertation are described as well as the methodologies, processes and techniques to be used.

In chapter 5 the results of the various different models used are presented and discussed.

Chapter 6 contains a conclusion to the thesis.

Chapter 2

Background

The purpose of this chapter is to highlight and explain important concepts that will be used throughout the following chapters and how they fit within the context of the thesis.

2.1 Demand Estimation and Forecasting

Demand estimation and forecasting is the process of estimating how much product that customers will purchase and thus, the activity of building forecasting models plays a key role in the management of a company's supply chain [2, 3]. Due to the increasing globalization of the distribution network that spreads out its main actors throughout the globe, the lead times associated with the decentralization of the distribution network are increased and that, coupled with the fairly irreducible manufacturing lead times and the volatility of customer demands makes the optimization of a company's supply chain a vital factor to its success [4].

In addition to the constraints mentioned above, the fashion market also has to deal with specific peculiarities such as the large number of items, the relatively short lifetimes of items, the substitution of items for each collection, the horizon required for forecasting due to the long lead times involved, the dependence of sales on fashion trends, the disaggregation of fashion products in various sizes and colours among other characteristics and the influence of several exogenous variable such as the weather, holidays, marketing actions, promotions and economic factors [1, 5, 4, 6, 7].

Due to all of the constraints stated above, it is imperative to choose the right methods to ensure an accurate and swift forecasting of demand.

2.1.1 Models for Demand Estimation and Forecasting

The models commonly used for demand estimation and forecasting can be loosely divided in time series based and causal models.

2.1.1.1 Time Series Based Models

Time series based models are perhaps the most commonly used methods in sales and demand forecasting and have been researched for decades. These traditional forecasting techniques have their foundation in statistics and include approaches such as the Holt-Winters model, Box & Jenkins model, regression models or autoregressive moving average (ARMA) models and their extensions such as ARIMA, SARIMA, SARIMAX or FARIMA [5, 8].

Despite the wide usage of time series methods in sales and demand forecasting, these methods aren't easily or efficiently implemented in the fashion sector because most of these methods require large historical data sets, are limited in their linear structure, require the operator to be quite experienced and versed in statistics but they also require an intricate parametric optimization [5]. Moreover, it can also be said that most of these traditional techniques and methods find their efficiency in seasonal or cyclical data, the demand for fashion products are affected by so many variables that traditional statistical methods may be rendered ineffective for forecasting demand [9, 10]. It can also be said that sales data can have a lot of outlier values and missing data (apart from the censored sales data) that must be cleaned and interpolated for time series approaches to be successful [1, 10]. Moreover, because fashion products present no continuity, it wouldn't be possible to use time series successfully.

These drawbacks clash quite clearly with the complexities of demand estimation and forecasting stated in 2.1.

2.1.1.2 Causal Models

In recent years, data mining models have emerged as an alternative to the more classic time series based methods and are becoming ever so present in many aspects of modern society [11] such as email spam filtering [12, 13, 14], text-to-speech transcriptions [15, 16], content recommendation systems [17] and image recognition [18, 19].

Machine learning, as a sub-field of soft computing and a branch of AI can be loosely divided in three major categories: supervised learning, unsupervised learning and reinforcement learning [20].

In supervised learning, arguably the most common category of machine learning, the training data given to the learning algorithms comes in the form of labelled examples that are fed into the algorithm, allowing it to make a prediction and afterwards giving it feedback with relation to the accuracy of the prediction. Over time, it is expected that the algorithm observes a new example from the testing data set and predicts a proper label for it [11, 20].

Within supervised learning, one could loosely divide supervised learning problems in *classification* problems and *regression* problems.

Classification is a typical supervised learning task, and can be defined as the task of assigning objects to one of several predefined categories. Regression is the other task commonly associated with supervised learning and is used to predict a target numeric value [20, 21]. Examples of classification tasks include the categorization of cells as benign or malignant based on MRI scans and the classification of galaxies based on their shapes, and some examples of regression include stock market index predictions and the projection of sales of a company based on the amount of money employed for advertising [21]. Some examples of commonly used ML algorithms for both classification and regression include, but are not limited to, k-nearest neighbours (KNN), Bayes classifiers, Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees, Random Forests, Linear Regression and Logistic Regression.

In addition to these methods, other methods called ensemble methods are commonly used. These types of methods use multiple models in order to obtain a better overall predictive performance than any of its constituent models [22]. Random Forests, for example, can be seen as an ensemble of decision trees (fig. 2.1).

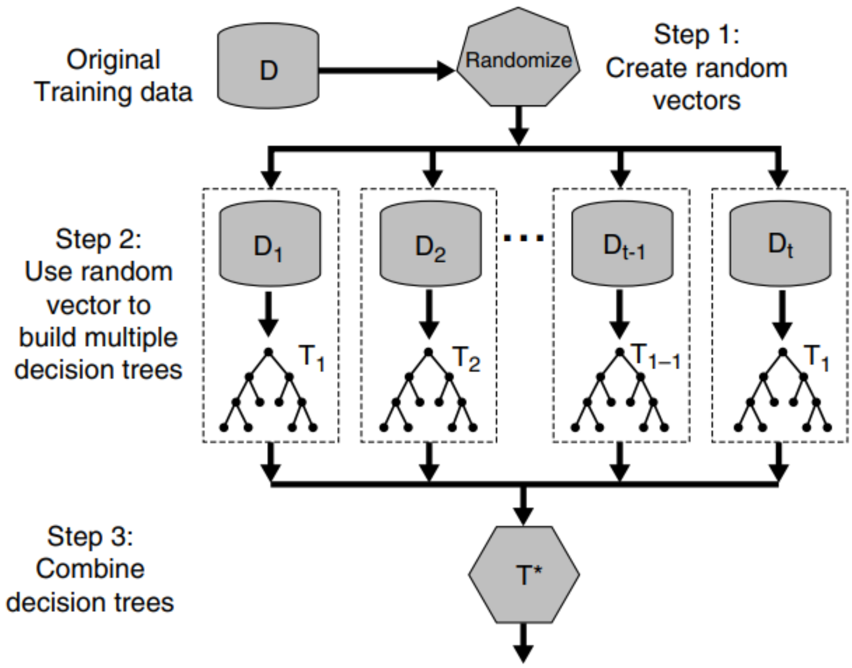


Figure 2.1: Architecture of a generic random forest model [21]

In unsupervised learning, the training data set is unlabeled, the data objects are grouped based only on information found in the data that describes the objects and their relationships [20, 21]. The task most commonly associated with unsupervised learning is *clustering*. Clustering (or cluster analysis) divides the data into different clusters that are meaningful, useful, or both [21]. Some common clustering methods include k-means clustering, hierarchical clustering, expectation-maximization and self-organized maps.

2.2 Summary

Some of the most commonly used methods for demand estimation and forecasting were briefly described in this section. In the following section, we are going to take a deeper look at how these methods are being used in similar fields of research and related works.

Chapter 3

Literature Review

The related works in the literature review will be analyzed according to a few points pertaining to the focus of research, the data being used with respect to the range and granularity in the product, location and time dimensions, the predictive variables chosen and the baseline methods being used.

3.1 Focus

Loureiro et al. [1] developed a forecasting model to estimate the sales data for products for which *no historical data* exists, the main focus being the exploration of deep learning approaches to forecast sales in the fashion industry by predicting the sales of new individual products in future seasons. This focus on forecasting without historical data is also pursued by Thomassey et al. [4, 6] where the authors forecast the short term sales profiles for numerous new items through a neural clustering and classification system [4] and a clustering / decision tree based system [6]. Choi et al. [23] have developed a forecasting algorithm, combining ELMs and the grey model, to forecast the sales of a knitwear fashion company, with limited data and time. Pavlyshenko [10] uses a stacking approach to study various models for sales forecasting in the presence of very little historical data (for example, when a new product or store is launched). Kaneko and Yada [24] construct a sales prediction model for retail stores using a deep learning approach using the machine learning framework H2O in R, focusing on the sparsity of sales data as well.

When *historical data* are available the goal is to extract as much information as possible from the past years, and for fashion products this information are trends and seasonality, but also the impact of exogenous factors which are where the greatest difficulty lies and where the most advanced techniques are most advantageous [5]. Wong and Guo [25] focused on tackling the medium-term fashion sales forecasting problem, using preprocessed historical sales data to generate forecasts using an HS–ELM learning algorithm, developed to improve NN generalization ability by integrating an HS algorithm (a metaheuristic) with an ELM (fig. 3.1).

Sun et al. [26] have also used an ELM to research the relationship between sales amount and various important factors affecting demand, using three sets of historical fashion sales data

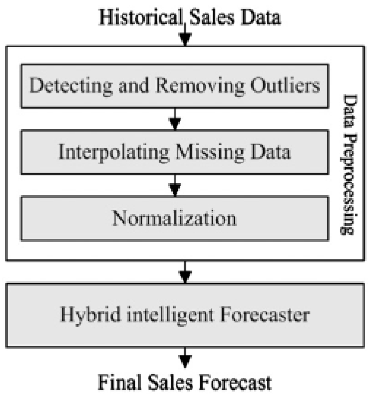


Figure 3.1: Framework of the hybrid intelligent model developed by Wong and Guo [25]

provided by a Hong Kong fashion retailer. Fuzzy Inference Systems have been used in numerous fields of research and in the fashion field, Thomassey at al. [27] have focused their research on short-term forecasts by readjusting forecasts from historical sales data, using an ANFIS based model. Ramos et al. [28] propose two forecasting systems based on ETS and ARIMA models in order to compare the performance between these two distinct methods, using historical data to predict the sales of footwear. Yu et al. [29] focus on combining ELMs and traditional statistical methods using historical data of several fashion Stock Keeping Units (SKU) from a fashion retailing company in Hong Kong. A slightly different approach, based on evolutionary computation has been proposed by Au et al. [30] for the forecasting of apparel sales using historical data obtained from a fashion retailer while in another instance of historical data usage, Lu [31] focuses on combining the variable selection method MARS (multivariable adaptive regression splines) and support vector regression (SVR) to construct a hybrid sales forecasting model for computer products. Islek and Oguducu [2] address the problem of demand forecasting for various products of main Turkish distribution warehouses through a hybrid forecasting model combining moving averages and a Bayesian Network. Yue et al. [32] research the problem of demand forecasting using support vector machines (SVM) and compare it to other models such as Winter's model and radial basis function networks (RBFNN).

3.2 Data

Loureiro et al. [1] used a data set referring to the sales of 684 types of women's bags, during the seasons Spring-Summer of 2015 and Spring-Summer of 2016 comprised of historical sales data for each product and its physical characteristics, logistical and internal aspects of the company and the opinion of domain experts, aggregated over all the stores of the entire chain (> 900 stores).

Thomassey et al. [4, 6] used data relative to a French textile distributor and the historical data, used for the learning process, are composed of 482 items corresponding to the years of 1998 and 1999, while the life span of all items was normalized to 52 weeks in order to compare and cluster sales profiles. Whereas the life span of items considered by Thomassey is relatively quite large (corresponding to 1 year), Loureiro at al. considered much shorter life spans, between 4 and 8 weeks, depending on the classification of the product with relation to the fashion and season's trends.

In order to study the performance of state space and ARIMA models for sales forecasting, Ramos et al. [28] analyzed the monthly sales of five categories of women's footwear in 70 stores of a Portuguese brand of footwear, between January 2007 and April 2012 and made 64 observations). These data contrast with the previous studies with respect to the number of items considered and the fact that Ramos et al. did not consider product life cycles at all.

Choi et al. [23] on the other hand, don't describe the data in a lot of detail, besides it being categorized in various types, several styles and colours, though one can assume that product life cycle is going to be very low as is the case with fast fashion products [33].

Pavlyshenko [10] used a big data set from Google's Kaggle community for store sales of a chain with over 3000 stores. The author experiments two forecasts for different time periods, one forecast for a long period (1 year) and another with a short period (3 days).

Kaneko and Yada [24] use three years of POS data collected between 2002 and 2004 in Japanese supermarkets. The model created by the authors was then used to predict an increase or decrease of sales for the next day.

Wong and Guo [25], who used historical data, research the medium-term sales forecasting problem through three experiments which make monthly, quarterly and annual forecasting. In each experiment they considered the sales amounts of 4 cities, deemed the 4 most important for the company's business, and 4 item categories, deemed the most influential on the company's business: skirts (spring/summer), jackets (spring/summer), coats (fall/winter) and pants (fall/winter). These categories being in different stations allow the authors to study the seasonality aspect as well.

Sun et al. [26] research the forecasting of sales for only one product (jeans), with 12 different colours, 51 different sizes and 182 different prices for a month. In the second experiment, the authors consider the sales data of socks (only one size), while in the third and final experiment data for jackets is studied, again divided by colour, size and price. The time granularity for this study is based on months and the forecasts are done for the time horizon on one year.

Thomassey et al. [27], for the forecasting model using a Takagi-Sugeno based ANFIS, used a database of 322 historic items from a textile distributor, with the available series length being three years (156 weeks). The authors created both mean-term and short-term forecasting models, with horizons of 1 week and 52 weeks, respectively. For the mean-term forecasting model, the authors used 2 years for the learning process and one year for evaluation while for the short-term forecasting model, the learning process is carried out on the first two years and the first weeks of the third year.

Like Choi et al. [23], Yu et al. [29] don't describe the time horizons for forecasts in detail, though the focus is fast sales forecasting so one would assume that the horizon would be low. The authors studied two data sets, one from a fashion retailing company based in Hong Kong and another from an online fashion shop. The first data set is composed of sales data of several fashion SKUs, as well as other properties related to the SKUs, such as the sales amount, colour, size and price of the product and the data set consists of 120 samples while the second data set consists of POS data of three months of sales from the shop's log.

Au et al. [30] on the other hand create a forecasting model from week 5 to week 8 of sales for two different products from a fashion retailer using sales data collected in the years 2002 and 2003 and for the first article, the life cycle is of around 70 days while the second article has a life cycle of around 8 months.

Lu [31], in order to evaluate the performance of the proposed forecasting model, used the weekly sales amount data of five computer products collected from a Taiwanese company for the period between January 2005 and September 2009, having 247 data points in the data set.

Islek and Oguducu [2] use the sales data of products from a Turkish company between the years of 2011 and 2013 from 98 warehouses and 70 products.

Yue et al. [32] used data pertaining to sales quantity of different brands of beer from a chinese retailer. In their research, sales quantity of 212 weeks have been generated.

3.3 Predictive Variables

With regards to the predictive variables and features being used, Loureiro et al. [1] use a set of 10 variables that characterize the products, with the price of the product being the only numerical variable, all others being categorical. The variables are then categorized in three classes, namely, product characteristics, logistical and internal organizational aspects of the company and domain experts. The authors also study the significance of the predictive variables, finding out that the most important are "Expectation Level" (a classification system to model the sales expectation of products), "Store Type" and "Family". Thomassey et al. [4, 6] emphasize the importance of the selected descriptive criteria on the influence on observed sales but assert that this choice is generally imposed by the availability of criteria in the retailer's database. In both [4] and [6], the criteria chosen were the price, starting time of sales and life span of items. The authors state, however, that additional criteria such as style or textile material would be of great interest for forecasting purpose but these data were not available in the database. Choi et al. [23] use variables such as the type of product, style and colour. In their application of the ELM for fashion sales forecasting, Sun et al. [26] have chosen variables such as colour, size and price and find that these factors have significant impacts on sales and like Sun et al., Yu et al. [29] have chosen the variables colour, size and price. For the ANFIS based forecasting system, Thomassey et al. [27] consider the selling price (promotion or no promotion), holidays periods and year sequences.

On the other hand, Ramos et al. [28] rely on sales data alone, not considering other characteristics for the products. Like Ramos et al., Wong and Guo [25] rely solely on sales data for their forecasting model as is also the case for Au et al. [30]. Lu [31] has a different approach with regards to the variables being chosen, using a MARS model to choose the most important variables.

Pavlyshenko [10], using a Random Forest algorithm was able to determine feature importance for various categorical features such as promo (whether a product was having a promotion or not), day of the week, day of the month and month. The results for feature importance show that the most important feature was whether a product was on sale or not.

In Kaneko and Yada's research [24] the sales data for each day were aggregated according to product attributes. The three categories considered range from one very broad category (considering 62 attributes) to a very specific category (considering 3312 attributes). For example, vegetable products would be described in category 1 as "Vegetables", in category 2 as "General Vegetables" and in category 3 as "Tomato" or "Cucumber", an example of hierarchical clustering.

Islek and Oguducu [2] consider attributes relating to warehouses and products. With relation to warehouses, the authors consider attributes such as location, number of customers and number of sub-warehouses. With relation to products, the authors consider product category, selling amount and selling time.

Yue et al. [32] have found that demand is not only relative to the historic sales data but also affected by factors such as whether there are active promotions or not for that week, seasonal factors, sales quantity for the previous week, average sales of the previous four weeks and beer prices for the current week.

Qi et al. [34] of a variety of features in their recurrent neural network based forecasting system inside five categories, namely, static features, date features, user behavior features, purchasing features and promotion features.

In summary, it can be observed that while many of these authors use similar types of predictive variables, not many of them take into account variables related to stock-out events, therefore possibly missing out on the potential of these types of events to improve forecasting accuracy.

3.4 Methods Used

Loureiro et al. [1] have tested a deep neural network (fig. 3.2) and compared it to several other methods, namely, decision trees, random forests, support vector regression, artificial neural networks and linear regression, having found that the deep neural networks and random forests had the best performance compared to the other methods tested.

Kaneko et al. [24] have also tested a deep neural network and compared it to logistic regression, finding out that for the same number of attributes, the deep learning model obtained a predictive accuracy around 10% better than logistic regression.

Thomassey et al. [4] propose a Neural Clustering and Classification model, based on clustering using a Self-Organizing Map Neural Network and classification using a Probabilistic Neural Network, finding out that the Neural Clustering and Classification model increases the accuracy of mid-term forecasting in comparison with the mean sales profile predictor. Thomassey et al. [6] also propose a model, based on existing clustering technique (k-means algorithm) and decision tree classifier (C4.5 algorithm), reaching the conclusion that the model is useful to estimate sales profiles of new items with no historical sales data and the model allowed an overall increase of the accuracy of mid-term forecasting in comparison with the mean sales profile predictor and other tested classifiers. Thomassey et al. [27] have also proposed an ANFIS based model for short-term

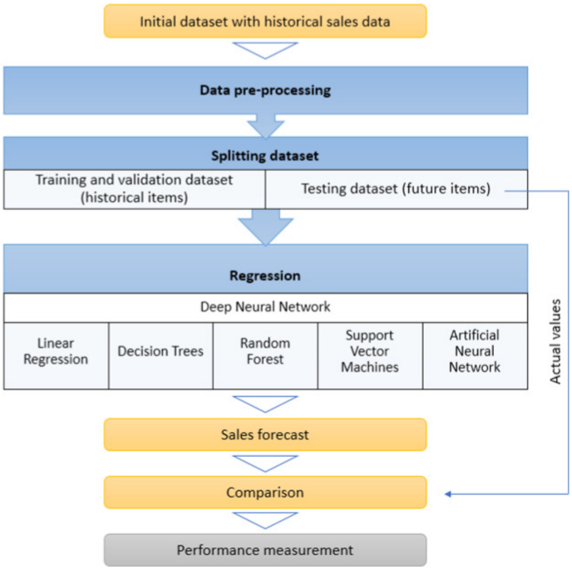


Figure 3.2: Schematic representation of the methodology followed by Loureiro et al. [1]

forecasting, having found out that it performs better than other methods, namely, Holt-Winters, ARMAX and naive models.

Choi et al. [23] have proposed a model combining an Extended ELM and grey model, coming to the conclusion that it performs better than either of them alone. Sun et al. [26] have proposed an extension of ELM (ELME) model with a regression integration method with the results demonstrating that the proposed methods outperform several sales forecasting methods based on backpropagation neural networks. Wong and Guo [25] have proposed a model based on an ELM and harmony search algorithm combination with the results demonstrating that the performance of the proposed model is much superior to traditional ARIMA models, evolutionary neural networks and the ELME model proposed by Sun et al. Yu et al. [29] propose a method combining ELM and traditional statistical methods with the results showing that this method outperforms traditional ANNs and statistical models.

Pavlyshenko proposes a model based on a stacking approach of different methods, including ANNs, ARIMA, Random Forests and LASSO regression, with the results showing that using stacking improves the performance of predictive models, and that stacking outperforms any of the other techniques used by themselves.

Ramos et al. [28] compare the forecasting performance of state space models and ARIMA models reaching the conclusion that the performance is similar on both one-step and multi-step forecasts.

Au et al. [30] propose an Evolutionary Neural Network (ENN) model and compare it to traditional forecasting approaches like SARIMA, with results showing that the ENN can produce comparable forecasting results to SARIMA and can even outperform it when the coefficient of variation of daily demand within a week does not have big variations and the demand does not exhibit persistent seasonal trend. Moreover, the authors find that the ENN approach for forecasting is a highly automatic one, while the SARIMA model involves more human knowledge

Lu proposes a combination of MARS and Support vector Regression to forecast sales, with the results showing that the proposed model can produce better prediction precision and outperforming the competing four models, namely, GA-SVR (genetic algorithm - suppport vector regression), ARIMA, pure MARS AND pure SVR. Yue et al. [32] have proposed a SVM based model, with the results showing that using a SVM improved the forecasting accuracy of the model, when compared to other methods such as traditional statistical models, Winter's model and RBFNN. The authors also affirm that the prediction accuracy of the SVM could also be improved if ensemble-learning techniques were to be used.

Islek and Oguducu [2] propose a hybrid model (moving average and bayesian network) with the results showing that the forecasting results are very encouraging if separate models were to be applied for sub distribution warehouse clusters instead of one model for all main warehouses or models for main distribution warehouse clusters.

Qi et al. [34] propose a novel for sales forecasting, based on Seq2Seq, a type of Encoder-Decoder model using recurrent neural networks (RNN), specifically the gated recurrent unit (GRU) variant of RNNs. On top of this decoder, the authors introduce a sales residual network to model the impact of the competing relation when a promotion is launched for a product The results show that the propose model obtains statistically significant improvements to the accuracy of forecasting with relation to other models such as AR, ARIMA, DNN and CNNs. The novelty of this research lies in the fact that sales are estimated in a recurrent fashion, with results showing substantial performance gain over traditional baselines and up-to-date deep learning alternatives.

In summary, it can be observed that many authors have used times series based models in the past and that causal models have seen an increase in popularity as of late but not many authors have explored deep learning models for demand forecasting in the fashion industry, so this is certainly a conceivable route to explore.

3.5 Summary

The relevant literature was analyzed with regards to the focus of research, the data, variables and methods being used. This literature review has brought to light the quantity of research using machine learning methods being researched, as opposed to the more traditional statistical methods. It has also brought to light that the majority of authors research sales forecasting and not demand forecasting.

Chapter 4

Data & Methodology

The objective of this chapter is to describe the problem at hand, as well as characterize the available data and identify possible solutions.

4.1 Scope

In the retail context, an erroneous determination of the amounts to buy of each article from the suppliers, either by excess or defect, can result in unnecessary costs of storage or lost sales, respectively. Both situations should be avoided by companies, which promotes the need to determine purchase quantities efficiently.

Currently companies collect huge amounts of data referring to their sales and products' features. In the past, that information was seldom analyzed and integrated in the decision making process. However, the increase of the information processing capacity has promoted the use of data analytics as a means to obtain knowledge and support decision makers in achieving better business outcomes.

This project aims at exploring the use of data mining techniques to optimize the amounts to buy of each product sold by a fashion retail company, resulting in the development of a model that uses past sales data of the products with similar characteristics to predict the quantity the company will potentially sell from the new products. The project uses as a case study a Portuguese fashion retail company.

4.2 Available Data

4.2.1 Product Data

The Portuguese fashion retail company used as a case study has over 900 stores scattered around the globe, with more incidence in the Iberian Peninsula. This company offers a wide range of products such as handbags, wallets, jewellery, watches and various other accessories, but the data used in this work are related to the sales of 708 women's bags SKUs during the homologous Spring-Summer 2015 and 2016 seasons.

The data included in the company's database can be divided in product data and sales data. With relation to product data, some of the variables included in the company's database are "Family", "Subfamily", "Colour Type", "Colour", "Matching", "Size", "Price", "Expectation Level", "Segment" and "International". "Family" identifies the material and/or pattern used in product production (fig. 4.1). "Subfamily" that identifies the format of the product. "Colour Type", identifies if the product is single colour or multi colour. "Matching" identifies if a product matches with another product from a different type of product (wallets for example).

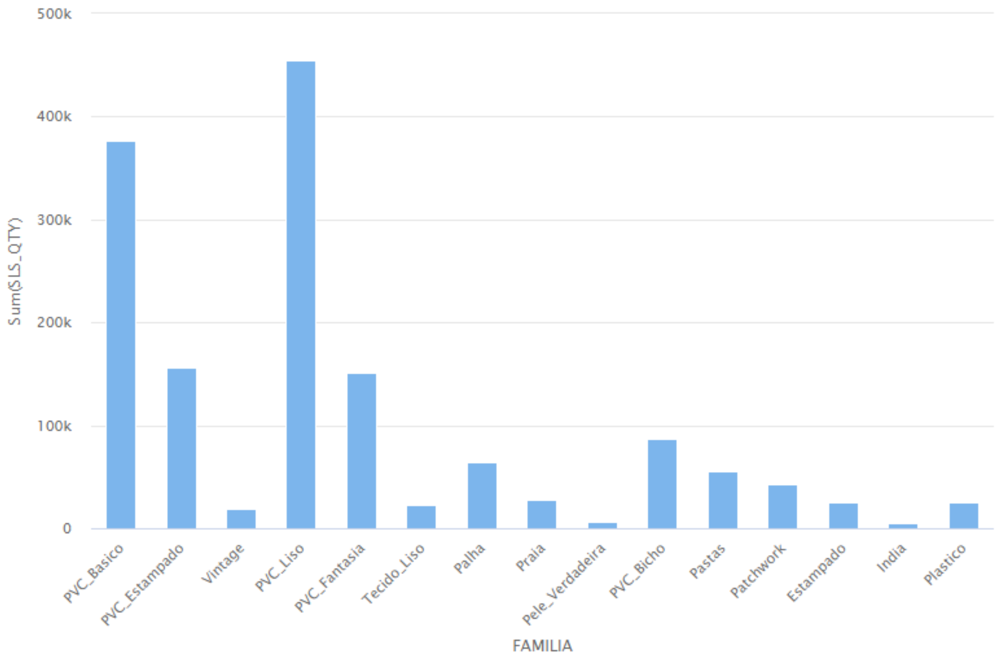


Figure 4.1: Total sales aggregated by product family for the 2015 Spring-Summer season

"Expectation Level" identifies the sales expectation for the product, ranging from 'M3' (low) to 'SB' (highest) (fig. 4.2). "Segment" identifies the target group ("Women" or "Teen"). "International" identifies whether or not a product can go to all markets of the brand or not.

The other two variables pertaining to products require a lengthier explanation. The variable "Store Type" (fig. 4.3) contains the types of store where the product will be available for sale. The company uses rates their stores from 'A' to 'D', 'A' being a store with a large display area and high sales potential that will receive all the products for a certain season and 'D' being a smaller store with a more limited variety of products due to its low sales potential.

The variable "Fashion" (fig. 4.4) is related to the season's trends and to the amount of time the product will be available in stores, classified according to the current season's trends from "Trendy" to "Basic", "Distribution Centralized" and "Basic Fashion". The articles classified at "Distribution Centralized" are articles similar to others from previous seasons and whose behaviour is expected to be similar due to their likeness. Furthermore, according to the "Fashion" type, the product is available in stores for different periods of time, namely, "Trendy" for 4 weeks, "Basic Fashion" for 6 weeks, "Basic" for 8 weeks and "Distribution Centralized" for 8 weeks.

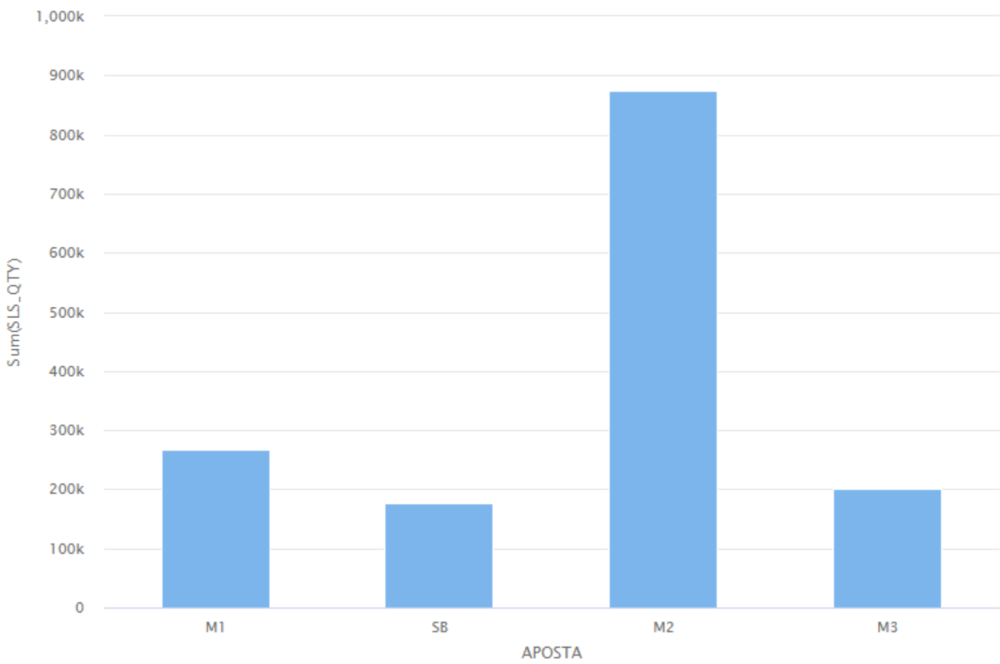


Figure 4.2: Total sales aggregated by product expectation level for the 2015 Spring-Summer season

4.2.2 Sales Data

With regards to sales data, the variables available in the company's database are the "Sales quantity" i.e. the total amount of units sold and "Stock quantity" i.e. the total amount of units in stock. A table summary of all the attributes used can be seen in table 4.1.

4.3 Methodologies

This project's objective is to use the data from the company's database to predict product demand. Keeping this objective in mind, the impact of different techniques and methods on prediction is analysed. In order to achieve this, various data mining methods to predict demand are tested and the results compared.

In order to automatically discover useful information in the dataset used in the project, the Knowledge Discovery in Databases (KDD) framework was used. Several steps are involved in the KDD process, such as data processing, data mining and postprocessing, as shown in fig. 4.5.

4.3.1 Data Preprocessing

This part of the process aims to make data more suitable for analysis. Firstly, the datasets pertaining to sales and products were inner-joined using the PROD_COD attribute. After that, the resulting dataset was split into testing, training and validation sets. Firstly, the data from the 2015 Spring-Summer season and the data from the 2016 Spring-Summer season were used for testing

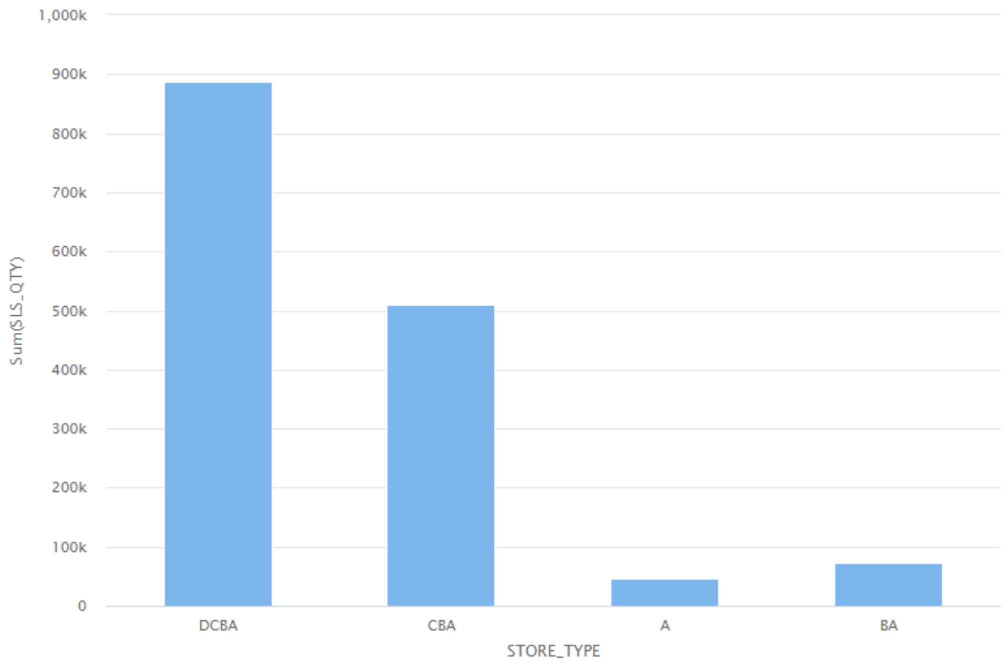


Figure 4.3: Total sales aggregated by store type for the 2015 Spring-Summer season

and training sets, respectively and then the training data were split 80/20 between the training subset and validation set, respectively.

4.3.2 Data Mining Techniques

At this stage of the process, the data are preprocessed and ready to be fed to the data mining algorithms to learn. The following data mining techniques were used:

-Linear regression;

-Decision trees;

-Random forests;

-Support vector machines;

-Neural networks;

-Deep neural networks

4.3.2.1 Linear regression

Regression is a set of various different statistical methods used to predict the value of one or many dependent variables from a series of independent variables. In the case of linear regression, the relation between one dependent variable and many independent variables is studied and the value of the dependent variable is predicted. In order to achieve this, a linear equation is fitted to the data.

This linear regression algorithm uses the Akaike information criterion in order to evaluate model quality and goodness of model fit. The formula for this criterion is as follows:

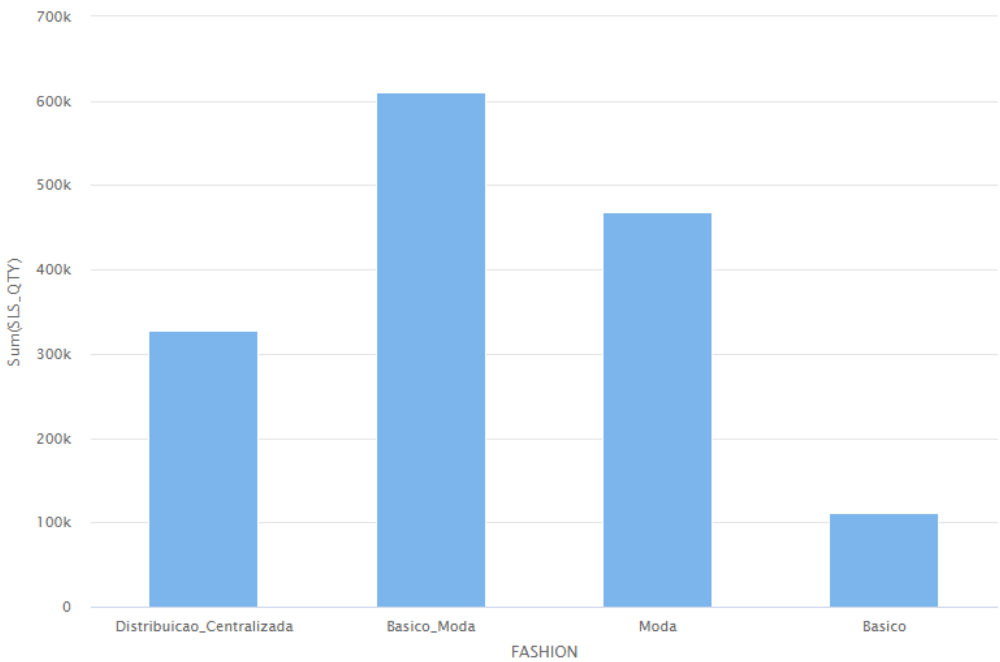


Figure 4.4: Total sales aggregated by product fashion level for the 2015 Spring-Summer season

$$AIC = 2k - 2\ln(\hat{A})$$

Where AIC represents the Akaike information criterion, k the number of estimated parameters in the model and $\hat{A}$ the maximum value of the likelihood function for the model.

The parameters optimized in the linear regression learning algorithm were the maximum iterations for feature selection, the minimum tolerance for eliminating collinear features and the feature selection method used during regression.

With each iteration of the optimization process, a model is obtained, applied and evaluated used the root mean squared error, absolute error, squared error and squared correlation criteria.

4.3.2.2 Decision trees

Decision trees are a simple and extensively used regression technique. In this technique, the regression process is modeled with the use of a set of hierarchical decisions on the feature variables, arranged in a tree-like structure [35].

A decision tree is composed by three types of nodes: a root node, internal nodes and leaf nodes. The leaf nodes are associated with class labels and the internal nodes contain test conditions to separate instances with different characteristics. The depth of a decision tree is the longest distance between the root node and a leaf node. A truncated example of a decision tree generated by the algorithm can be seen in figure 4.6.

In the decision trees learning algorithm, the parameters optimized were maximal depth, minimal leaf size and the criterion used was least square. The maximal depth parameter is used to

Name	Attribute	Type
Expectation level	APOSTA	Nominal
Color	COLOR	Nominal
Days in exposition	DAY_EXPOS	Numerical
Day of the week	DAY_WEEK	Nominal
Family	FAMILIA	Nominal
Fashion	FASHION	Nominal
Store code	LOC_COD	Nominal
Product code	PROD_COD	Nominal
Price	PREÇO_BASE_IVA	Numerical
Segment	SEGMENTO	Nominal
Sales quantity	SLS_QTY	Numerical
Stock quantity	STK_QTY	Numerical
Store type	STORE_TYPE	Nominal
Subfamily	SUBFAMILIA	Nominal
Size	TAMANHO	Nominal
Color type	TIPO_COR	Nominal
Week of the year	WEEK_YEAR	Numerical

Table 4.1: Attributes in the company's database

restrict the maximal depth of the decision tree, the minimal leaf size refers to the minimal number of examples in its subset and the criterion refers to the criterion used to choose which attributes will be used for splitting nodes.

4.3.2.3 Random forests

Random forests are simply an ensemble of decision trees, hence the name forest. The trees from this random forest are created and trained by feeding a set to the algorithm and by using bootstrapping subsets from the entire dataset.

For the random forests learning algorithm, the parameters optimized were maximal depth and subset ratio. The subset ratio criterion serves to specify the ratio of random attributes to test.

4.3.2.4 Support vector machines

The support vector machine is a classification technique that works well with high-dimensional data and avoids the dimensionality problem. This technique is also able to represent the boundary of decisions, known as the support vectors, using a subset of training data [21].

SVMs work by transforming the original data into a higher dimension and in that dimension searching the optimal decision boundary and optimal hyperplane. When data is properly mapped into a higher dimension, it's always possible to separate data from two classes [36]. SVMs are less prone to overfitting than other methods and can also be used for both classification and numeric predictions.

In the support vector machines learning algorithm, the parameters optimized were C and kernel gamma.

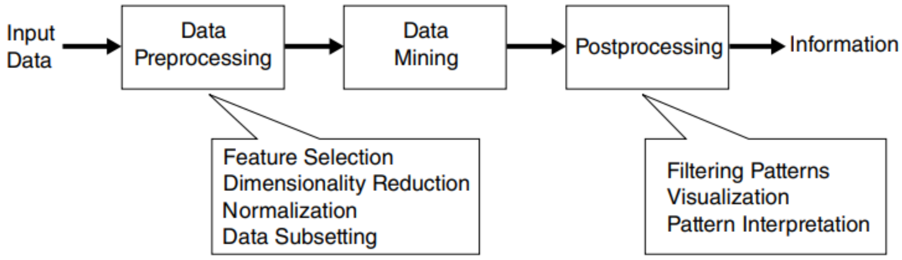


Figure 4.5: The process of knowledge discovery in databases (KDD) [21]

4.3.2.5 Neural networks

Neural networks are computational simulations of human neurons. In humans, the nervous system is composed of interlinked neurons. These neurons are connected by synapses, the links between axons and dendrites, and learning occurs by changing these synaptic links between neurons. An analogy between biological neural networks and artificial neural networks can be seen in figure 4.7.

The simplest ANN model is the perceptron. A perceptron consists of two node types, input nodes and an output node. The input nodes represent the input attributes and the output node represents the output of the model. In order to connect these nodes, several weighted links are established. This link simulates the synaptic links between biological neurons. The objective of training a perceptron is then to modify the links' weights until they model the relationships of the data. The perceptron obtains the output value by computing the weighted sum of all its inputs, subtracting the bias and looking at the sign of the result. The output can be generalized as follows [21]:

$$\hat{y} = sign(w_d x_d + w_{d-1} x_{d-1} + ... + w_2 x_2 + w_1 x_1 - t)$$

Where $\hat{y}$ is the output value of the perceptron, $w_1, w_2, ..., w_d$ are the weights of the links, $x_1, x_2, ..., x_d$ are the input values and t is the bias factor. The sign function acts as the activation function and outputs 1 if the result is positive and -1 if the result if negative. In figure 4.8 an example of a modelled boolean function can be seen.

In a perceptron, the weight parameters w are trained according to the training examples and adjusted until the outputs match the outputs of the training examples. The formula involved in the learning algorithm is the following [21]:

$$w_j{}^{(k+1)} = w_j{}^{(k)} + \lambda (y_i - \hat{y}_i^{(k)}) x_{ij}$$

Where $w^{(k)}$ corresponds to the weight parameter relating to the i^{th} input after the k^{th} iteration, x_{ij} is the value of the j^{th} attribute relating to the training example x and λ is the learning rate. Some authors also add a momentum parameter in order to improve speed of learning [38].

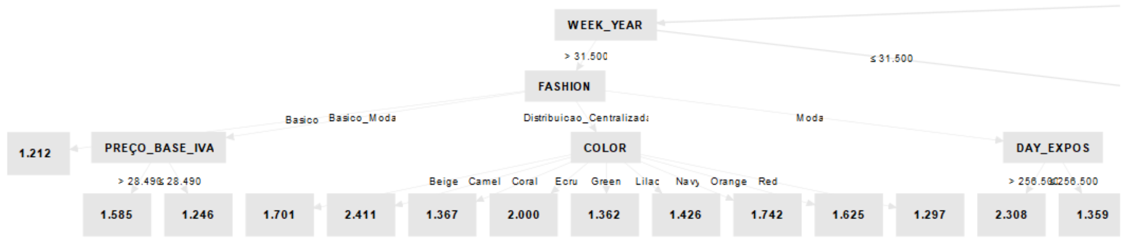


Figure 4.6: Fragment of a decision tree generated by the training algorithm

A more complex type of neural network is the multilayer artificial neural network. These types of neural network use intermediary layers and nodes called respectively hidden layers and hidden nodes. Another important definition is feedforward. In feedforward neural networks, the nodes in each layer are only connected to nodes in the next layer. In figure 4.9 we can see an example of a feedforward neural network with one hidden layer, similar to the one used in the project.

The parameters optimized in the neural networks learning algorithm were the learning rate and momentum. The neural network architecture includes a hidden layer with ten hidden nodes.

4.3.2.6 Deep neural network

The deep neural network architecture used in the project is based on a feedforward artificial neural network with multiple layers and it is trained using gradient descent and backpropagation. Gradient descent or method of steepest descent is an iterative optimization algorithm widely used in machine learning to find local minima and maxima of a function.

Like in the perceptron case, other neural networks use activation functions, albeit more complex ones, in order to process the incoming signal. The usage of nonlinear activation functions is important in order to introduce nonlinearities into the network [39].

For the deep neural network algorithm, the parameters optimized were the epochs, learning rate and activation functions. The activation functions tested were tanh, ReLU, ELU and Maxout.

4.3.3 Model evaluation

In order to train the models and then evaluate their performance, a number of ways of measuring error such as the mean squared error (MSE), root mean square error (RMSE), absolute error (MAE) and squared correlation were used.

The MSE takes the difference between the observed values y_i and predicted values $\hat{y}_i$, squares it, sums it and then divides it by the number of data points, as such:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

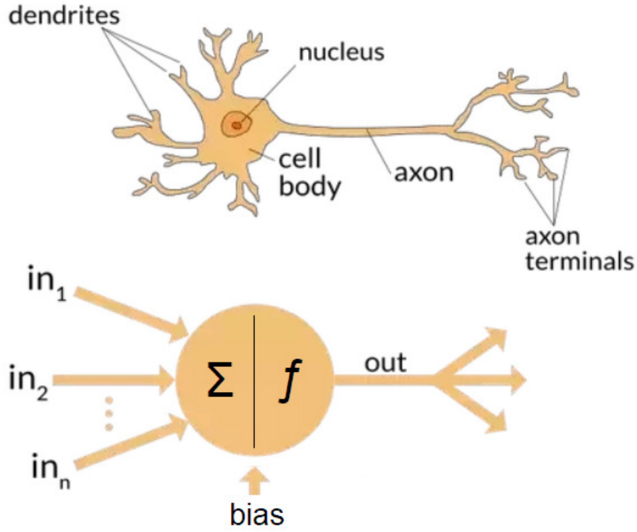


Figure 4.7: Analogy between biological neural networks and artificial neural networks [37]

The root mean squared error is the root of the MSE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

The mean absolute error is the average of the absolute values of the errors, as such:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

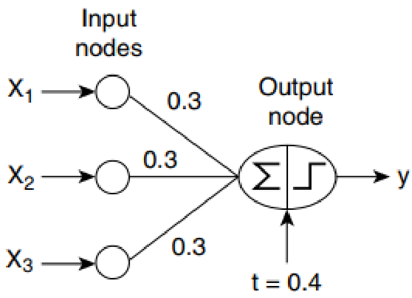
The squared correlation, r^2, is the square of the correlation coefficient:

$$r^2 = \left(\frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}\right)^2$$

,

X_1	X_2	X_3	y
1	0	0	–1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	–1
0	1	0	–1
0	1	1	1
0	0	0	–1

(a) Data set.



(b) Perceptron.

Figure 4.8: Example of a modelled boolean function using a perceptron [21]

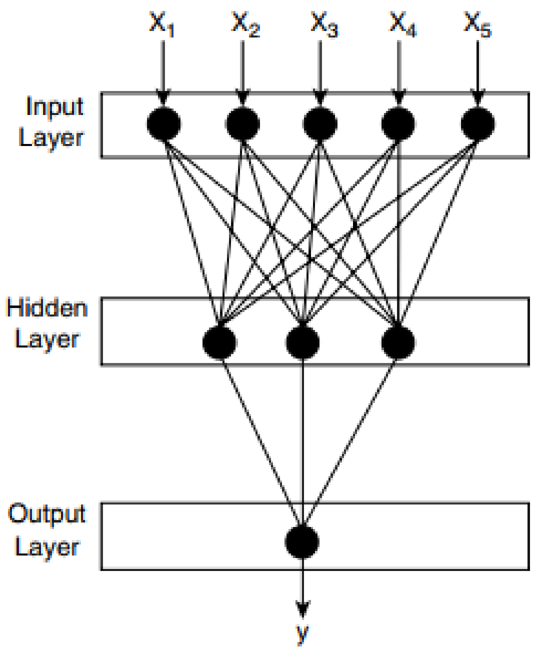


Figure 4.9: Example of feedforward ANN with one hidden layer [21]

Chapter 5

Results & Discussion

In this section, the results of the various different models used and how they compare to the baseline model using only an aggregated average for predictions are presented and discussed. A short summary of these results can be seen in table 5.1, with the best results for each evaluation metric presented in bold.

	MSE	*RMSE*	*MAE*	r^2
Baseline	26266560.810	5125.091	3325.482	0.578
Linear regression	7326489.760	2706.749	1472.266	0.415
Decision trees	3802204.262	1949.924	1300.369	0.686
Neural network	4303909.809	2074.587	1139.602	0.687
Support vector machine	4663848.501	2159.595	1315.551	0.738
Random forest	3822888.192	1955.221	1193.004	**0.786**
Deep neural network	**3495308.155**	**1869.574**	**1062.939**	0.770

Table 5.1: Performance results for the various regression techniques on the testing dataset

The technique with better overall performance was the deep neural network, with the random forest coming close and even surpassing it in the r^2 metric. Surprisingly decision trees had a good performance as well for their simplicity. Other regression techniques had worse performance results in the evaluation metrics, despite some techniques, like the support vector machine with its r^2 of 0.738 and both neural network techniques coming close to these techniques and every model performed better than the baseline one, except for the linear regression in the r^2 metric.

The results from table 5.1 show that no single technique can be considered the best, as no technique achieved better results in all error metrics. Although the deep neural network achieved the best result in three error metrics, it still came in second place relative to the random forest in the r^2 metric. If r^2 is considered the main error metric, than random forests can be considered the best performing technique, but if not then the deep neural network would have to be considered the best performing technique.

The results also show that deep neural networks can also be used for relatively small data sets and that simple techniques such as decision trees can also be reliably used for demand prediction.

From the point of view of the company, the selection for best technique should consider an equilibrium between simplicity, performance and intelligibility. In this regard, a deep neural network would be more complex to tune for the average operator than a random forest, which is more intuitive in its approach and still providing excellent results.

In fig. 5.1 a scatter plot of observed sales against predicted sales for the best performing model can be seen. In the plot we can see a positive correlation that matches the 0.786 value in the r^2 metric.

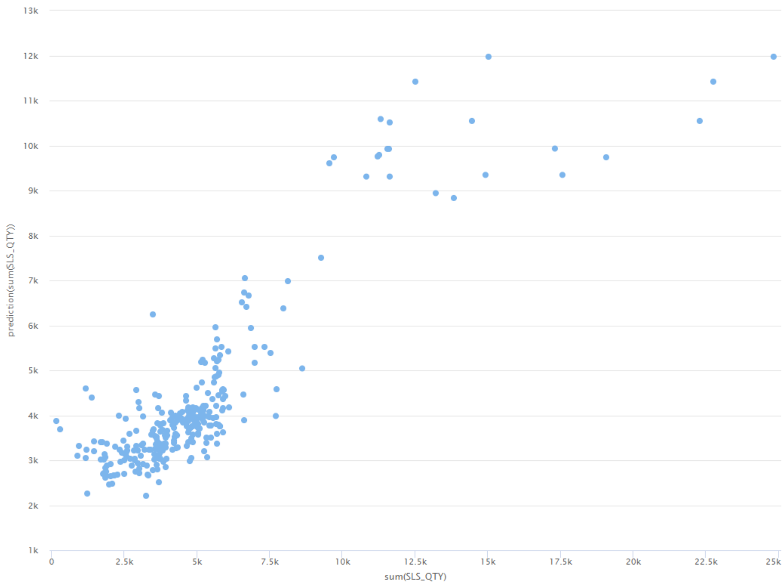


Figure 5.1: Scatter plot of observed sales and predicted sales for the random forest

It can also be seen in fig. 5.2 that the deep neural network also has a strong positive correlation, which matches its 0.770 r^2 score.

The support vector machine (fig. 5.3) also has an observable positive correlation albeit not as strong as the random forest and deep neural network, which matches its lower r^2 of 0.738.

Like in the support vector machine case, there is an observable positive correlation between observed and predicted sales for the decision tree (fig. 5.4) but not as strong as the deep neural network and random forest, which matches its r^2 of 0.686.

It should be noted that throughout all these techniques the model had some varying difficulties predicting higher values, mostly where the sum of sales is above 10k, and this can be particularly seen in the neural network and linear regression models (figs. 5.5 and 5.6).

Overall, it can be said that these results show that these techniques and the models produced with them, can be used as aids to predict demand and ultimately aid in support the company in decisions regarding marketing campaigns and quantities of stocks of each product to buy with reliability.

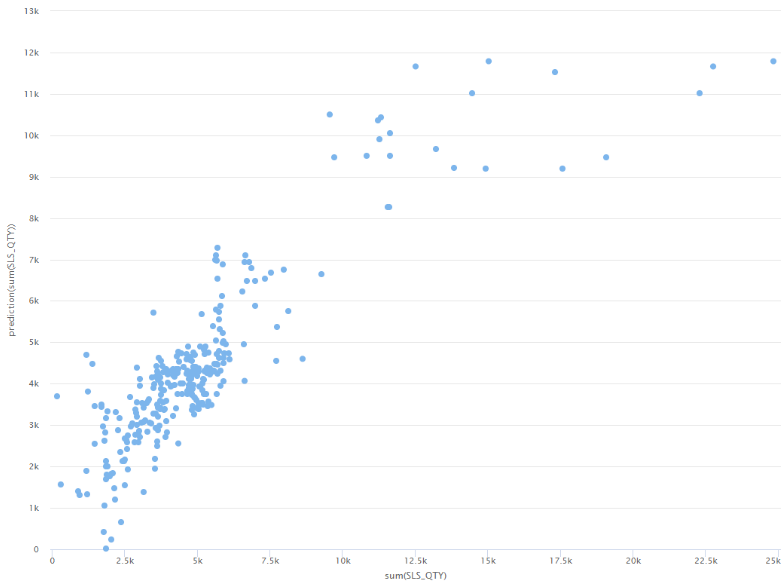


Figure 5.2: Scatter plot of observed sales and predicted sales for the deep neural network

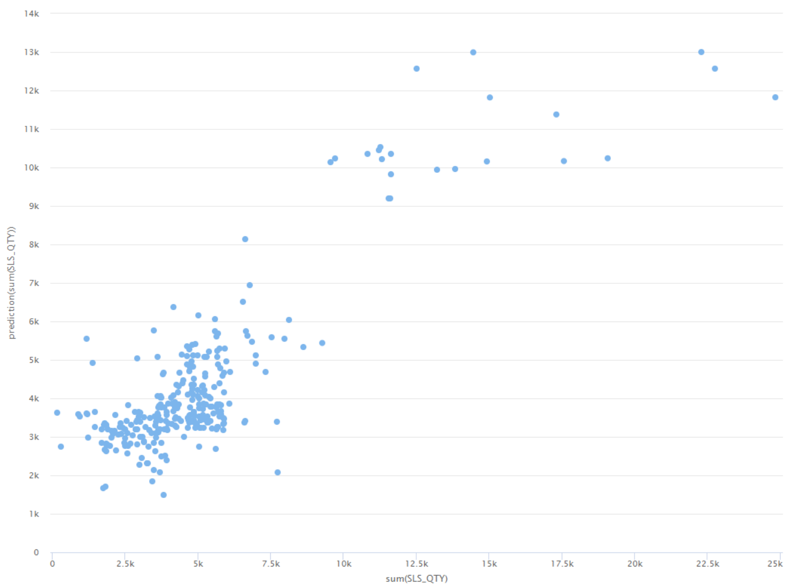


Figure 5.3: Scatter plot of observed sales and predicted sales for the support vector machine

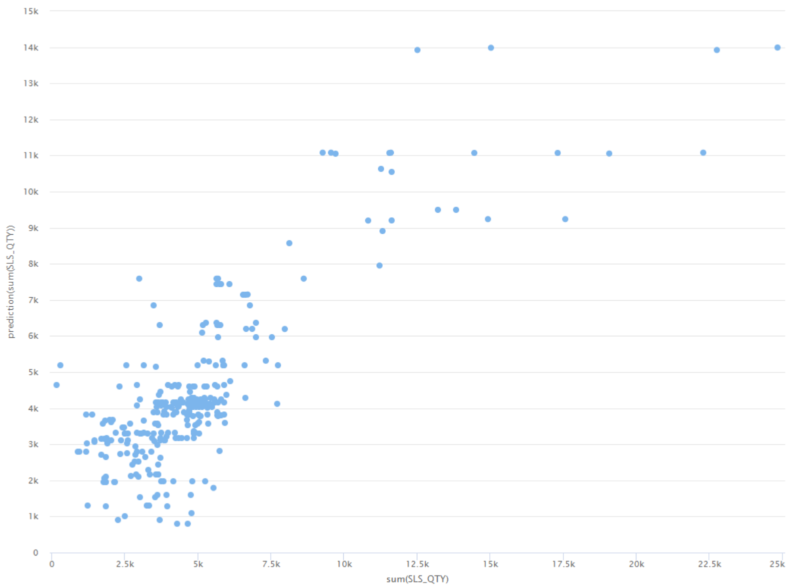


Figure 5.4: Scatter plot of observed sales and predicted sales for the decision tree

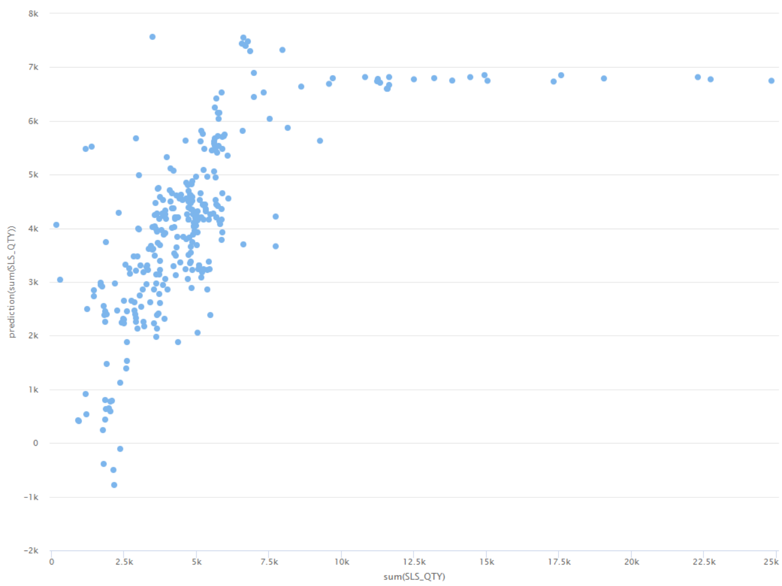


Figure 5.5: Scatter plot of observed sales and predicted sales for the neural network

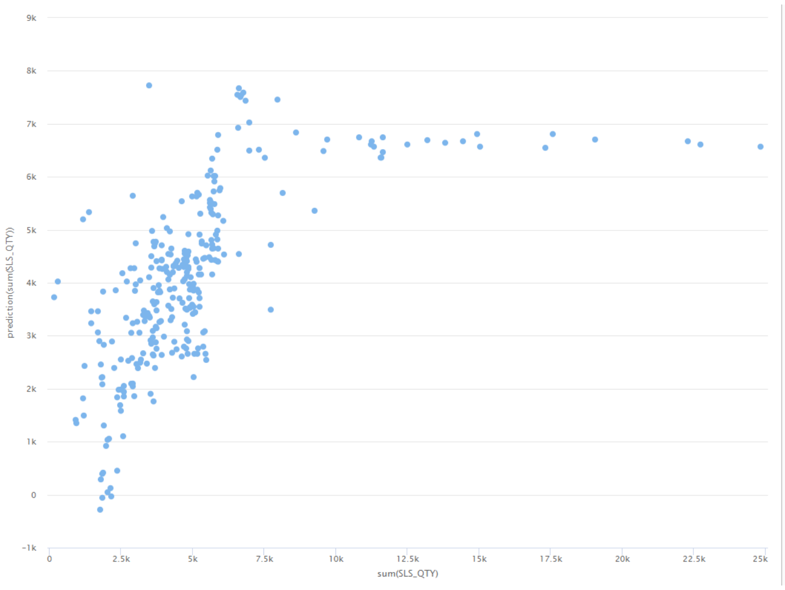


Figure 5.6: Scatter plot of observed sales and predicted sales for the linear regression

Chapter 6

Conclusions

The objective of this project was to develop a predictive model for demand forecasting. The data set includes data from over 900 stores of a large Portuguese fashion retail chain relating to the sales of 708 women's bags during the homologous Spring-Summer 2015 and 2016 seasons.

Several different regression models were created using various different techniques such as linear regression, decision trees, support vector machines, neural networks, random forests and deep neural networks. The data from the Spring-Summer 2015 season was used as training data and the data from the Spring-Summer 2015 season was using as testing data.

The results from the created regression models show that the deep neural network model and the random forest model had the best predictive results, with the deep neural network model achieving an r^2 of 0.770 and the random forest model achieving an r^2 of 0.786.

With these results, it can be concluded that after the creation and training of the model, it could be used by the company to reliably estimate the quantities of product to order for the Spring-Summer season of the following year.

Regarding future work, it would be interesting to explore demand prediction in situation where stockout events occur, as well as exploring new techniques and new sources of data that can be used to extract more predictive variables in order to increase model accuracy and performance.

References

[1] A. Loureiro, V. L. Miguéis, and L. F. da Silva, "Exploring the use of deep neural networks for sales forecasting in fashion retail," *Decision Support Systems*, vol. 114, pp. 81–93, oct 2018.

[2] I. Islek and S. G. Oguducu, "A retail demand forecasting model based on data mining techniques," in *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)*, IEEE, June 2015.

[3] A. Chawla, A. Singh, A. Lamba, N. Gangwani, and U. Soni, "Demand forecasting using artificial neural networks—a case study of american retail corporation," in *Advances in Intelligent Systems and Computing*, pp. 79–89, Springer Singapore, Sept. 2018.

[4] S. Thomassey and M. Happiette, "A neural clustering and classification system for sales forecasting of new apparel items," *Applied Soft Computing*, vol. 7, pp. 1177–1187, Aug. 2007.

[5] S. Thomassey, "Sales forecasting in apparel and fashion industry: A review," in *Intelligent Fashion Forecasting Systems: Models and Applications*, pp. 9–27, Springer Berlin Heidelberg, Oct. 2013.

[6] S. Thomassey and A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees," *Decision Support Systems*, vol. 42, pp. 408–421, Oct. 2006.

[7] S. Thomassey, "Sales forecasts in clothing industry: The key success factor of the supply chain management," *International Journal of Production Economics*, vol. 128, pp. 470–483, Dec. 2010.

[8] J. Wang, G. Q. Liu, and L. Liu, "A selection of advanced technologies for demand forecasting in the retail industry," in *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, IEEE, Mar. 2019.

[9] H. Huang and Q. Liu, "Intelligent retail forecasting system for new clothing products considering stock-out," *Fibres and Textiles in Eastern Europe*, vol. 25, no. 1, pp. 10–16, 2017.

[10] B. Pavlyshenko, "Machine-learning models for sales time series forecasting," *Data*, vol. 4, p. 15, Jan. 2019.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[12] Y. Cao, X. Liao, and Y. Li, "An e-mail filtering approach using neural network," in *Lecture Notes in Computer Science*, pp. 688–694, Springer Berlin Heidelberg, 2004.

[13] E. M. Bahgat, S. Rady, and W. Gad, “An e-mail filtering approach using classification techniques,” in *Advances in Intelligent Systems and Computing*, pp. 321–331, Springer International Publishing, Nov. 2015.

[14] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, “Machine learning for email spam filtering: review, approaches and open research problems,” *Heliyon*, vol. 5, p. e01802, June 2019.

[15] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” *arXiv:1609.03499*, 2016.

[16] Y. Ning, S. He, Z. Wu, C. Xing, and L.-J. Zhang, “A review of deep learning based speech synthesis,” *Applied Sciences*, vol. 9, p. 4050, Sept. 2019.

[17] S. Zhang, L. Yao, A. Sun, and Y. Tay, “Deep learning based recommender system: A survey and new perspectives,” *ACM Computing Surveys*, vol. 52, pp. 1–38, Feb. 2019.

[18] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, “Handwritten digit recognition with a back-propagation network,” in *Advances in neural information processing systems*, pp. 396–404, 1990.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, pp. 84–90, May 2017.

[20] M. Kang and N. J. Jameson, “Machine learning: Fundamentals,” in *Prognostics and Health Management of Electronics*, pp. 85–109, John Wiley and Sons Ltd, Aug. 2018.

[21] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson, 2006.

[22] I. Witten, E. Frank, and M. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems, Elsevier Science, 2011.

[23] T.-M. Choi, C.-L. Hui, N. Liu, S.-F. Ng, and Y. Yu, “Fast fashion sales forecasting with limited data and time,” *Decision Support Systems*, vol. 59, pp. 84–92, Mar. 2014.

[24] Y. Kaneko and K. Yada, “A deep learning approach for the prediction of retail store sales,” in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, IEEE, Dec. 2016.

[25] W. K. Wong and Z. X. Guo, “A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm,” *International Journal of Production Economics*, vol. 128, pp. 614–624, Dec. 2010.

[26] Z.-L. Sun, T.-M. Choi, K.-F. Au, and Y. Yu, “Sales forecasting using extreme learning machine with applications in fashion retailing,” *Decision Support Systems*, vol. 46, pp. 411–419, Dec. 2008.

[27] S. Thomassey, M. Happiette, and J. M. Castelain, “A short and mean-term automatic forecasting system—application to textile logistics,” *European Journal of Operational Research*, vol. 161, pp. 275–284, Feb. 2005.

[28] P. Ramos, N. Santos, and R. Rebelo, “Performance of state space and ARIMA models for consumer retail sales forecasting,” *Robotics and Computer-Integrated Manufacturing*, vol. 34, pp. 151–163, Aug. 2015.

[29] Y. Yu, T.-M. Choi, and C.-L. Hui, “An intelligent fast sales forecasting model for fashion products,” *Expert Systems with Applications*, vol. 38, pp. 7373–7379, June 2011.

[30] K.-F. Au, T.-M. Choi, and Y. Yu, “Fashion retail forecasting by evolutionary neural networks,” *International Journal of Production Economics*, vol. 114, pp. 615–630, Aug. 2008.

[31] C.-J. Lu, “Sales forecasting of computer products based on variable selection scheme and support vector regression,” *Neurocomputing*, vol. 128, pp. 491–499, Mar. 2014.

[32] L. Yue, Y. Yafeng, G. Junjun, and T. Chongli, “Demand forecasting by using support vector machine,” in *Third International Conference on Natural Computation (ICNC 2007)*, IEEE, 2007.

[33] T.-M. Choi and S. Ren, “Intelligent demand forecasting supported risk management systems for fast fashion inventory management,” in *Information Systems for the Fashion and Apparel Industry*, pp. 263–271, Elsevier, 2016.

[34] Y. Qi, C. Li, H. Deng, M. Cai, Y. Qi, and Y. Deng, “A deep neural framework for sales forecasting in e-commerce,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management - CIKM '19*, ACM Press, 2019.

[35] C. C. Aggarwal, *Data mining: The textbook*. Springer, 2015.

[36] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2011.

[37] R. Booto Tokime and X. Maldague, “Automatic defect detection for x-ray inspection: Identifying defects with deep convolutional network,” in *Conference: Canadian Institute for Non-destructive Evaluation (CINDE)*, 2019.

[38] N. Qian, “On the momentum term in gradient descent learning algorithms,” *Neural Networks*, vol. 12, pp. 145–151, Jan. 1999.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.