Graphical user interface

Description automatically generated with medium confidence

A Deep Learning System for Predicting Size and fit for clothing with Height to weight distribution - fit analytics Collaboration

with E -commerce company Saenguin (SNGN Suits GmbH)

*Master-Thesis*

Fakultät für Information, Medien und Design

SRH Hochschule

Heidelberg

Master’s in Applied Computer Science

submitted to

Supervisor: Prof. Christoph Hanh

External Advisor: Ms.Svenja

Submitted by: onkar Phatangare

Student ID: 11013949

Address: SRH Campus, Bonhoeffer straße 13, 69123 Heidelberg

Phone Number: +49 15901347796

E-Mail Address: omkar15september@gmail.com

Submission Date:

# Acknowledgment

The successful outcome of the thesis would not have been possible without the involvement, support, and superior guidance of many brilliant personalities. I would like to thank my research supervisor, Prof. and my colleague at company, Frau and for their continuous and unconditional support throughout the research period. Without their assistance and dedicated involvement in every stage of the process the paper would never be accomplished. I also appreciate the support from my colleagues and who have directly or indirectly supported me for the research by discussing their views, Interviews, and for providing me with quality organizational data.

**Abstract**

Researchers have been studying different methods to effectively predict the Useful prediction systems allow traders to get better insights about data such as: future trends. Also, investors have a major benefit since the analysis give future conditions of the market. One such method is to use machine learning algorithms for forecasting. This project’s objective is to improve the quality of output of stock market predicted by using stock value. A number of researchers have come up with various ways to solve this problem, mainly there are traditional methods so far, such as artificial neural network is a way to get hidden patterns and classify the data which is used in predicting stock market. This project proposes a different method for prognosting stock market prices. It does not fit the data to a speciﬁc model; rather we are identifying the latent dynamics existing in the data using machine learning architectures. In this work we use Machine learning architectures Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN) and Hybrid approach of LSTM + CNN for the price forecasting of NSE listed companies and differentiating their performance. On a long-term basis, sling window approach has been applied and the performance was assessed by using root mean square error.

**Acknowledgments**

This thesis is based on research work This work would not be possible without many people whose contributions can’t be ignored.

for providing required resources for this work. I wish to express my sincere appreciation to my supervisors whose assistance was a milestone in completion this project. Prof persistently guided me in planning of work and solved my queries. I owe my gratitude to my classmate who helped in understanding different machine learning techniques. Without their support, a significant portion of project work was not possible in constrained time.

I must also appreciate my family and friends for helping me survive all the stress throughout the year. To my parents, for supporting me both on and off the water. I would like to thank you all mentioned and other people who have helped me directly or indirectly for pushing me farther than I thought I could go.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **SR.**  **NO.** | **TOPIC** | **PAGE**  **NO.** |
|  | Abstract…………………………………………………………… | I |
| Acknowledgements……………………………………………….. | II |
| Table of Contents…………………………………………………. | III |
| List of Figures…………………………………………………….. | IV |
| List of Table………………………………………………………. | VI |
| **1.** | **Introduction………………………………………………………** | **1** |
|  | 1.1 Prediction……………………………………………………… | 1 |
|  | 1.2 Application …………………………………………………… | 3 |
|  | 1.3 Objectives……………………………………………………... | 3 |
|  | 1.4 Motivation…………………………………………………...... | 3 |
|  | 1.5 Organization of Report…………………………….………….. | 3 |
| **2.** | **Literature Survey…………………………………………………** | **4** |
| **3.** | **Existing work…….…………………....…………………………..** | **10** |
|  | 3.1 Overview of Existing work….………………………………… | 10 |
|  | 3.2 Proposed work…………………………………………………. | 11 |
|  | 3.2 1 Working of model……………………………….…………… | 12 |
|  | 3.2.2 Working of model……..……………………….……………. | 14 |
|  | 3.2.3 Working of approach model….…........................................... | 15 |
| **4.** | **Dataset, Implementation and Results…………………….……..** | **16** |
|  | 4.1 Dataset Details………………………………………………… | 16 |
|  | 4.2 Tools & Technology used…………………………………….. | 17 |
|  | 4.3 Implementation Results……………..……………………..….. | 20 |
| **5.** | **Conclusion and Future work.…………………..…..……………** | **30** |

[Bibliography and References 31](#_TOC_250004)

[APPENDIX A: ABBRAVIATION… 34](#_TOC_250003)

|  |  |  |
| --- | --- | --- |
|  | **LIST OF FIGURES** |  |
| **FIGURE NO** | **TITLE** | **PAGE NO** |
| 3.1 | Proposed Work Flow | 11 |
| 3.2 | LSTM Architecture | 12 |
| 3.3 | CNN Architecture | 14 |
| 4.1 | Stock Dataset Information | 20 |
| 4.2 | Read Dataset | 20 |
| 4.3 | Graph of Close Price history | 21 |
| 4.4 | Data Scaling | 21 |
| 4.5. | LSTM Summary | 22 |
| 4.6 | CNN Summary | 22 |
| 4.7 | Hybride Approach of LSTM + CNN Summary | 23 |
| 4.8 | Training Process | 23 |
| 4.9 | Predicted Close Price | 24 |
| 4.10 | Plot for Real vs Predicted value for ICICI Bank using LSTM | 25 |
| 4.11 | Plot for Real vs Predicted value for ICICI Bank using CNN | 25 |
| 4.12 | Plot for Real vs Predicted value for ICICI Bank using LSTM + CNN | 25 |
| 4.13 | Plot for Real vs Predicted value for SUN PHARMA using LSTM | 26 |
| 4.14 | Plot for Real vs Predicted value for SUN PHARMA using CNN | 26 |
| 4.15 | Plot for Real vs Predicted value for SUN PHARMA using LSTM + CNN | 26 |
| 4.16 | Plot for Real vs Predicted value for GSFC using LSTM | 27 |
| 4.17 | Plot for Real vs Predicted value for GSFC using CNN | 27 |
| 4.18 | Plot for Real vs Predicted value for GSFC using LSTM + CNN | 27 |
| 4.19 | Plot for Real vs Predicted value for RS SOFTWARE using LSTM | 28 |
| 4.20 | Plot for Real vs Predicted value for RS SOFTWARE using CNN | 28 |

|  |  |  |
| --- | --- | --- |
| 4.21 | Plot for Real vs Predicted value for RS SOFTWARE  using LSTM + CNN | 28 |
| 4.22 | Plot for Real vs Predicted value for VARDMANPOLY using LSTM | 29 |
| 4.23 | Plot for Real vs Predicted value for VARDMANPOLY using CNN | 29 |
| 4.24 | Plot for Real vs Predicted value for VARDMANPOLY using LSTM + CNN | 29 |

|  |  |  |
| --- | --- | --- |
|  | **LIST OF TABLES** |  |
| **TABLE NO** | **TITLE** | **PAGE NO** |
| 4.1 | Dataset Details | 16 |
| 4.2 | Accuracy | 24 |

**Chapter 1. Introduction**

### Predicting Size and fit for clothing with Height to weight distribution

Who does not adore the convenience and ease of shopping online? It is advantageous, offers a wide variety of products, and offers reasonable deals. In addition, retailers benefit from the Internet, which enables them to establish their brand. As a result, the internet-based fashion industry has experienced tremendous growth recently. As a result, looking for garments online is an extremely interesting and difficult process. This is due to wide variations in sizing between different brands that can make it difficult for customers to identify appropriate fitting clothing. There are two major consequences of this: first, it leads to a bad shopping experience for customers, as they will be required to return the items, and second, this results in money-related problems for retailers. Therefore, giving precise and customized fit guidance is essential for further improving the online shopping experience and lowering product return rates by enhancing further the online shopping experience.

Generally, when it comes to determining the correct size of clothing, customers can choose to determine their size just by examining their body and measuring it manually, or by choosing a size that they are accustomed to wearing. It is important to keep a few things in mind as you take this manual measurement. In most cases, the process is carried out with the use of a measuring tape and involves measuring your height, the width of your shoulders, the circumference of your waist, and the height of your hips. While these measurements may seem accurate and suitable for all types of clothing, they are not always accurate. Additionally, customers may also need to take additional measurements if they are purchasing clothing that is more specific (e.g., suits or dresses with long sleeves).

Therefore, in this study, we aim to develop a predicting system that will allow users to predict their body sizes (e.g., small, medium, large) for a perfect fit, by providing their body measurements (e.g., height, weight, age, body shapes, etc.) with proper height to weight distribution.

Gradually over the course of the past few years, some prediction models have been developed for this purpose and some of them have been applied to predicting clothing sizes. This classification is based on the following criteria:

* + 1. Statistical Analysis
    2. Comparative Analysis

#### Size-specific Analysis

#### Analysis of Body Measurement

#### Statistical Analysis

In order to handle gender differences in relation to age and BMI, all statistical analyses were stratified by analyzing both gender and age separately, as shown in previous research (Merrill et al., 2019). It is notable that male and female participants demonstrated similar torso BSP parameter trends regardless of age and BMI, and yet opposite trends for thigh COM locations as a function of BMI were observed (Merrill et al., 2019). Before any further analysis, the fifteen segment parameters of interest (mass, COM, and RG for the torso, thigh, shank, upper arm, and forearm) were tested for normality, to which they were then log-transformed as needed before further analysis. There were 280 participants in the full data set, so the data were randomly divided into two subsets: the training set, which had 200 participants, and the testing set, which had 80 participants. In analyzing the torso structure, thigh structure, limb structure, and upper arm/forearm segment parameters in the training subset, multiple regression analysis was performed with a backward elimination method for variable selection and stratified based on gender. As an initial step, the first models included age, body mass index (BMI), age and BMI interaction terms, waist, hip, and neck circumferences, and all relevant measurements taken of the body segment of interest. Every time the analysis was done, the predictor with the largest p-value was removed, allowing the analysis to be repeated for all the predictors. It was repeated until each remaining predictor's p values were below 0.10 with respect to the previous predictors. This study was carried out with the help of SAS Institute's JMP Pro 12 (Cary, NC, United States). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6905426/

In addition to the waist, hip, and neck circumferences being not directly measured, these measurements were incorporated to all initial models due to their benefit for determining overall body shape and mass distribution, specifically their ability to define central adiposity, which when combined with BMI can help describe the relative distribution of mass throughout the torso and appendages of individuals at varying levels of obesity. For instance, individuals with people with more central adiposity will have higher waist and/or hip circumferences than individuals with less central adiposity, which means that at the same BMI, individuals with bigger circumferences will have greater total and normalized limb mass, together with COM and Rg values that are identical to those of individuals with less obesity.

In the independent validation data set, models were used to predict in vivo segmental parameters which were compared against actual segments (selected by DXA in vivo). The error percentages and root mean square errors were also calculated, in order to compare the levels of error between the two data sets. It is also reported that the models explained the total variability within the data set when applied to the testing set (R2), along with the improvement of the models (\*R2) compared to previous models using age and body mass index only (Merrill, et al., 2019). As an additional effort, the actual values of the testing set were compared to a widely used segment parameter prediction method (de Leva, 1996) using the same metrics as the actual values .

#### Comparative analysis

It has been determined that, in this study, contrasts between body measurements obtained using the 2D direct body measuring method and those obtained using the 3D body scanning method were analyzed. Additionally, the fit of pants to which body measurements obtained by the two different measuring methods were applied was compared cross-sectionally. In this study, 10 females in their 20s- 30s were measured using direct body measuring, as well as automatic 3D body measurements, using Hamamatsu body scanners. Three of the women then wore experimental garments in order to test whether their measurements were accurate or not. In order to make experimental pants, they used their 2D direct body measurements and 3D automatic measurements, and they conducted wearing tests with the help of expert evaluations and cross-sectional evaluations. This research has been looking at the design of experimental pants. Based on the results of a comparative analysis of the differences in 2D direct body measurements and 3D scan measurements, the 3D automatic measurements were significantly larger in bust circumferences, ankle circumferences, armscye circumferences, shoulder lengths, depths of armpits, and arm lengths than the 2D direct body measurements. The circumference measured by the 3D body scanner was found to be slightly larger than that measured directly. Therefore, it is recommended that if the pattern making method used was easy, it should also be adjusted. A similar pattern making method was used to prepare experimental garments by applying body measurements obtained through two different methods, and those garments were evaluated through expert evaluations and cross-sectional 3D scan evaluations on the same pattern made. In the studies, it became apparent that 2D-pants using 2D direct measurements of the body on the waist, hip circumference, and abdominal circumference were slightly looser than 3D pants using 3D measurements. When comparing the results of the experiments by comparing the quality of fit and appearance of the experimental garments in terms of how the garments look on each of the subjects, significant differences were found in most of the comparisons. Hence, this outcome suggests that 3D automatic body measurement data for different body shapes would yield different accuracy results, hence it will be necessary to compare with 2D direct body measurements for different body shapes and 3D automatic measurements for different body shapes.

https://www.researchgate.net/publication/263649852\_Comparative\_Analysis\_of\_Body\_Measurement\_and\_Fit\_Evaluation\_between\_2D\_Direct\_Body\_Measuring\_and\_3D\_Body\_Scan\_Measuring

**Size-specific Analysis**

It is demonstrated in this study that analyzing body scan data of a company's target market in order to determine size-specific sizing can provide information that can be used to improve the fit of ready-to-wear apparel. An explanation is given as to how various size-related statistical and visual analysis methods can be used to gain insight into target market body scan data. In this way, it becomes possible to describe and address the range of body shapes and measurements that are present in a sizing system and identify potential design changes that can be implemented in order to increase the percent of acceptable fit within each size category for a target market.

https://textiles.ncsu.edu/tatm/wp-content/uploads/sites/4/2017/11/Loker\_full\_136\_05.pdf

**An analysis of the body's measurements**

In order to take the correct measurements of a person's body, the appropriate type of tape measure must be used. Use a soft cloth or a flexible plastic or rubber tape measure if possible, such as something that can be used for sewing. Therefore, the body measurement is carried out in order to ensure the results are accurate.When taking body measurements, you need to use the correct type of tape measure. If possible use a soft cloth or flexible plastic/rubber tape measure, such as is used in sewing . Accordingly , the body measurement is carried out to determined the accurate results.

### Applications

* Business
* Companies
* E commerce company
* Fashion Industry

### Objectives

### It is important to provide accurate and personalized fit recommendations in order to enhance the online shopping experience and reduce the amount of returns.

### Motivation

By analyzing the feedback customers provide on the purchased child products, companies can learn customers' sizes preferences and how products are made, allowing them to recommend better fitting sizes to customers. When it comes to the e-commerce clothing industry, where there are countless retailers, offering something unique will help retailers stand out from the competition.

Rationale of the study:

Machine Learning Algorithms

Almost every platform we are comfortable with at this point relies on machine learning - Google, Facebook, Siri, and many others. As a matter of fact, machine learning algorithms have played a very important role in the recent development of digital technology. A primary characteristic of contemporary artificial intelligence is its ability to find and predict patterns in a vast amount of data by using machine learning algorithms. These algorithms are capable of processing vast quantities of data including words, clicks and in general, anything that can be measured in a digital form.

A COMPETITIVE ADVANTAGE

Benefits of Machine Learning in Retail

Machine learning technology allows businesses to collect customer data in order to:

* Analyze data and make informed decisions
* Improve the accuracy of marketing plans
* Improve the quality of personalized services
* Enhance customer loyalty by providing a positive shopping experience
* Enhance sales

In order to remain competitive, apparel retailers strive to improve their customer relations, boost sales, and differentiate themselves from their competitors.Retailers with machine learning (and their powerful algorithms) are more likely to achieve these business objectives.Algorithms applied to fashion design. Toward a more sustainable apparel industry with machine learningRefund rates, sometimes referred to as "eCommerce's plague," are on the rise in the fashion industry. These actions are not only detrimental to the environment, but also place a heavy burden on the budgets and logistics of companies. However, AI and Machine Learning technologies are likely to lead to a reduction in return rates, resulting in a more sustainable industry.

There is a serious threat to the sustainability of the fashion industry posed by returns

By 2024, e-commerce is expected to have a market share of 60.32 percent, up from 46.6% in 2020. A whopping forty percent of online purchases are returned, compared with 8% to 10% of those purchased in stores.

In the e-commerce environment, consumers behave very differently than in brick and mortar stores, according to Prof. Ohnemus. It is common for individuals to buy clothes for special occasions and then return them, purchase items only to try them on, without intending to keep them, or order multiple sizes and colors in order to find the best fit.

"Many times, companies are unable to inform you what will be done with the products that you have returned. Despite being in perfect condition, most of these items end up in a landfill because the processing costs are too high for the company to bear. It is after they have been shipped for delivery in multiple countries or even across the globe." - Isabelle Ohnemus, founder and CEO, EyeFitU

The apparel industry could benefit from machine learning algorithms

Because one of the main reasons for returning clothes is that they are the wrong size or fit, with AI and machine learning specialists to create a software that can help customers identify the best possible garment for their body type and preferences.

Using machine learning and user-generated data, the algorithms are able to perform multi-parameter algorithms on one of the largest databases of body measurements in order to determine the right clothing size for a person, taking into account various body types and sizes. Its can be reported a 55 percent reduction in returns and a significant increase in conversion rates and average order values after implementation.

<https://www>.supertrends.com/algorithms-for-fashion-how-machine-learning-makes-the-apparel-industry-more-sustainable/

How Virtual Fitting Technology Works

In the past, the "Try before you buy" strategy was an effective means of attracting customers to apparel stores. Virtual fitting rooms now serve the same purpose. A report by Fortune Business Insights predicted that the virtual fitting room market size would reach USD 10 billion by 2027.

For a better understanding of virtual fitting room technology, let us examine the following example. A while back, a project for the development of augmented reality (AR) footwear fitting rooms. It is based on the following principle:

As a result of the input video being split into frames, a deep learning model is used to estimate the position of the key points associated with specific legs and feet.

To display the orientation to a user naturally, we place a 3D model of the footwear in accordance with the keypoints detected.

The three-dimensional footwear model is rendered so that each frame displays realistic lighting and textures.

ARKit is being used for the estimation of 3D human body poses and rendering of 3D models

In working with ARKit (a framework for augmented reality for Apple's devices), was found that the rendering is limited. the tracking accuracy is too low to be of any use for positioning footwear. Possibly, the cause of this limitation is due to the fact that inference speed is maintained while the tracking accuracy is neglected, which might be an issue for real-time applications.Furthermore, the ARKit algorithm performed poorly in identifying body parts. It does not detect any keypoints if only a part of the body is present in the processed image, because the algorithm is designed to identify the entire body. The algorithm is designed to process only the legs of the individual in a footwear fitting room.In the research, it was concluded that virtual fitting room applications may require additional functionality in addition to the standard AR libraries. The decision should be made to involve data scientists in developing a model that is capable of detecting key points on one or two feet in the frame and operating in real-time.

VIRTUAL TRY-ON CLOTHES

Comparatively to shoes, masks, glasses, and watches, virtual fittings for 3D clothing remain a challenge. As a result, clothes are deformed when they conform to the shape of a person's body. Consequently, for an effective AR experience, a deep learning model is required to recognize not only keypoints on the joints of the human body, but also the body shape in three dimensions.

In examining one of the most recent deep learning models, DensePose, which attempts to map RGB pixels of a person to 3D surfaces of the human body, it can be seen that it is not really suitable for augmented reality. For the fitting of 3D clothing items, DensePose's inference speed is insufficient for real-time applications, and its body mesh detection is not accurate enough. In order to improve results, more annotated data must be collected, which necessitates a substantial amount of time and resources.

Alternately, 2D clothing items and 2D silhouettes of people may be used. such a technology named Zeekit provides users with the opportunity to apply a variety of clothing types (dresses, trousers, shirts, etc.) to their images.

Applications and examples of machine learning in e-commerce (clothing industry):

Reuters reported in 2018 that Zara, an apparel retailer in Spain, was integrating artificial intelligence (AI) into its business strategy and supply chain in order to keep up with and improve upon its competition. It may be a coincidence, but Inditex, Zara's parent company, experienced strong profit growth the following year. Although 2020 - the year the COVID-19 pandemic shut down the world - was a tough time for everyone, so far Inditex has exceeded pre-pandemic levels for sales.

It is important to note that Zara is not the only fashion retailer utilizing AI and machine learning to remain competitive. There are many other brands engaged in the same practice, some of which are fast fashion and others haute couture, such as Tommy Hilfiger, H&M, and Dior. Recently, the apparel and accessories retailer GAP acquired the AI startup Context-Based 4 (CB4), which it is hoping will improve its customer experience through enhanced predictive analytics and demand forecasting.

According to a Juniper Research study published in 2021, 96% of surveyed retail executives intend to invest in artificial intelligence within the next three years. It's even more interesting to note that a survey conducted by NTT DATA and Oxford Economics reported that 40% of executives believe that AI is crucial to business success; they believe that failing to adopt the technology will negatively impact customers and employees as well as hurt the company's bottom line.

Retailers who view these statistics may feel compelled to consider using AI and machine learning in order to improve their business operations - and with good reason. Whether it is trend predictions, personalised recommendations, or automated customer services, fashion machine learning applications are numerous and varied. The following is a brief list.

In light of this, it is not surprising that an increasing number of fashion companies are making use of artificial intelligence and machine learning to enhance their processes:

Artificial intelligence-driven demand forecasting reduces forecasting errors by up to 50%.

In recent years, "AI designers" have been developed that have been able to analyze popular styles, design elements, including cut and fabric, as well as their past retail performance and popularity on social media, in order to create completely new designs.

The application of artificial intelligence (AI) to garment manufacturing improves productivity and improves time-to-market due to automated quality control and defect detection.

https://www.intelistyle.com/fashion-machine-learning-applications-and-examples/

Also, recommender systems and AI-based personalised styling are being implemented to enhance customer experiences during online shopping. Similarly, AI is playing an increasingly important role in the E-commerce and online shopping domains. AI recommends other similar items based on your color preferences, budget and other characteristics when you browse or search fashion items on e-commerce sites.

Most aspects of the fashion industry are being transformed by technological advancements, from the initial sketches through to fashion shows and to online shopping.

By using automation and predictive modeling, the automotive and fashion industries will be able to provide better customer service experiences for consumers, while reducing waste in the fashion industry through the use of digital technology.

Research objectives :

We are aware that clothing products come in a variety of sizes. Many online retailers nowadays allow customers to provide fit information on the purchased product (e.g. Small, Fit, or Large) during the returns process or when leaving a review.

Consequently, the goal of this research is to develop a predicting system that allows users to accurately predict their body shapes (e.g., small, medium, large) for the perfect fit by providing their body measurements (e.g., height, weight, age, body shapes, etc.) with a proper height to weight distribution.

Approach (Predictive model) :

Using deep learning models :

1. Linear regression

2. Random forest - classification, regression

3. Scikit-learn- Predictive Analysis

4. Body measurment data set

• Problem targeting - To what extent did the implementation resolve the research questions and objectives

• Duration - How long did it take to complete the project

• Utilization of resources - The implementation achieved the best use of the available resources, or did it result in non-relevant additions?

Architecture - How clean are the implementations of the programming language, frameworks, and data sources?

• Complexity - Does the application perform unnecessary operations, resulting in additional processing

• User Interface/User Experience - Are the application's actions unnecessary and causing excessive processing?

In machine learning, it is looking for patterns in the data, and based on those patterns, make predictions in order to answer business questions, detect trends and solve problems.

In retail, recommendation engines provide an excellent method of generating targeted and personalized marketing campaigns based on previously acquired information and predicted future behavior. If a customer purchases a black shirt, then it is very likely that the customer will also purchase a matching jean jacket. A retailer in this situation might wish to suggest to a customer who has purchased or looked at the black shirt that he or she also considers the jean jacket by saying, "customers who bought this black shirt also purchased this jean jacket.".

Change the words

As a result of this type of approach, fashion retailers can provide personalized recommendations for individual users rather than using more blanket manual or rule-based approaches to target clothing advertisements based on demographics. Thus, retailers' advertisements are much more likely to be impactful in generating sales, and thus they can use their advertising budgets more efficiently and reduce returns by suggesting items that users are more likely to be satisfied with.

The use of recommendation systems has proven to be one of the most widely adopted methods of filtering information. The simplest method of retrieval and recommendation is also used in the fashion industry. Users are solely responsible for selecting appropriate products based on their preferences. As a reminder, recommendation systems only suggest new trends and clothing to the user, but do not also take into account fashion experts who have a more in-depth understanding of different styles, color contrasts, or even knowledge obtained from having undergone a dressing course. In contrast, an automatic fashion composition system based on deep learning considers all of the above characteristics. Using deep learning as an approach to outfit composition, the system encodes visual features using a deep convolutional network which takes in a style outfit and predicts the user's engagement level. Based on the appearance and other metadata of the outfit candidate, the outfit is compiled by scoring the outfit. Using a multimode Deep Learning framework, the system becomes context-aware by taking advantage of the context from within the image. Multiple model approaches have been found to be better than single model approaches. An outline of a scoring model for an outfit that will be composed and recommended to a user is shown below.

Over the last few years, e-commerce has experienced an exponential growth in fashion. Customers expect a personalized experience, accompanied by suggestions for articles that are relevant to them. This is the key challenge for market players. Although product recommendations are a well-established field, advice regarding the size and fit of clothing is still relatively new. Data on the size and fit topic are extremely sparse and noisy, which makes the problem very challenging. Traditional machine learning approaches have been the most frequently employed until now. The aim of this work is to present a meta-learning approach that uses a deep neural network as a basis. This approach possesses the advantage of being able to exploit large amounts of data, to improve learning across categories of fashion, and to adapt to new data without retraining. This research compares the proposed approach to three recent methods that have been successfully applied in the domain, demonstrating its strengths and weaknesses. To this end, we utilize a large-scale anonymized dataset of customer-size interactions.

Research Questions:

In this study ,the main aim is to predict the actual body measurements to predict and provide accurate size recommendation measurement to design the clothes according to the fit .

* + 1. What are the main values which will be used to predict the measurement?
    2. How many values are given at the input to model the algorithm?
    3. How the model behaves with the input given values?
    4. What are the best possible ways to use the model?
    5. What model is used for the prediction is it follow regression or neural network approach?
    6. How accurate the model predicts the outcome results?
    7. What values the data set will consist for the prediction?
    8. What values the outcome will predict?
    9. What technology will be addressed?

An e-commerce platform for fashion must provide its customers with personalized fit and size recommendations. Predicting the right fit has a significant impact on customer satisfaction. Furthermore, it can reduce the costs associated with size-related returns.

### Organization of Report

Chapter 2 provides a review of the literature that summarizes the contributions of individual papers.

Chapter 3 describes the existing work on machine learning and how it is applied in the clothing industry, and some applications-based approaches.

Chapter 4 provides details on the implementation, the tools and technologies used, and the datasets.

In chapter 5, there are conclusions about machine learning and fit and measurement prediction, as well as future work about what you would like to pursue in the future.

**Chapter 2. Literature Review**

#### FITME: BODY MEASUREMENT ESTIMATIONS USING MACHINE LEARNING METHOD:

#### Publication Year: 2019

**Author:** Sahar Ashmawia , Maram Alharbia , Ameerah Almaghrabia , Areej Alhothalia

**Conference :** 16th International Learning & Technology Conference 2019

**Summary:** Abstract Online shopping platforms have attracted many customers since their introduction in the last decade of the 20th century. With the help of online shopping platforms, customers can purchase goods anytime and anywhere without having to go from store to store to find a product or wait in line at the checkout. Despite the advantages over shopping in a store, customers often have concerns when buying products where measurements need to be estimated, such as furniture and clothing. In particular, choosing the wrong clothing size is a common problem faced by many online shoppers. Therefore, in this study, we proposed a model that estimates people's body measurements from real-time images using Haar Cascade Classifier and Support Vector Machines.

Introduction Shopping is the activity in which customers look at the products or goods presented by some retailers with the intention of buying them. Shopping is one of the most important necessities of modern life, in which people buy things that meet their needs and interests. This process can take various forms, and online shopping is one of them. Especially in recent years, online shopping has taken over the world at a rapid pace. Customers prefer online shopping to in-shop shopping as it requires less time and effort. Customers with online shopping.

Related work : can search for any product in many stores and store online with just a few steps, rather than going into a store to find the products they want and standing in line at the checkout. Despite the many advantages of online shopping, some customers may encounter difficulties because they cannot judge the quality of the product, which may result in receiving a wrong, damaged or delayed product. Especially when shopping for clothes, problems of different kinds occur. There are many reasons why they prefer to store in a store rather than online, such as not being able to determine their correct size, try on the clothes, and assess the quality of the material. When determining the right size, customers either determine their clothing size by measuring their body manually or they choose a size they are used to wearing. The manual measurement is usually obtained using a measuring tape and physicallymeasuring the height, The shoulder width, bust, waist, and hip area. However, these measurements are not always accurate and suitable for all types of clothing. Customers may also need to measure additional body parts if they are purchasing more specific clothing (e.g., suits or long-sleeved dresses). For example, if a customer wants to buy a dress or jacket, he or she will need to measure the chest, waist, and hips. For example, to measure the waist area, the customer must measure the waist tightly and completely around the body. When measuring the thigh, the customer must measure the largest part of the thigh. After measuring the body parts, the customer must convert the measurements and choose the right size. Differences in measurements, body parts measured, and type of clothing often result in inaccurate manual measurements and cost time. Several studies have investigated the possibility of automatically estimating body measurements from customer images. Most of these approaches use specialized equipment (e.g., a depth camera) to capture 3D images of the human body. Despite the measurement accuracy that 3D-based methods provide, these methods are not suitable for all users who want to estimate their size while shopping online on the go. Few studies have used 2D images to estimate body measurements. Studies that rely on 2D images estimate some basic human body measurements or classify the human body into some predefined classes. Therefore, in this study, we aim to develop a smartphone system that allows users to estimate their body measurements and predict their body size by taking a 2D image of the body with a typical smartphone camera in a certain direction (e.g., front view). To develop a robust application and train the algorithm with real-world data, we first conducted an experiment by taking photos of a set of participants along with their body measurements. Then, the proposed approach extracts the features from the images using computer vision and machine learning techniques to estimate the body measurements (e.g., waist and bust size). Then, the approach uses a Support Vector Machine (SVM) to determine the appropriate size of the shoppers. The use of such a system will help online shopping customers accurately estimate their body measurements and improve the online shopping experience.

3. Methodology To estimate people's body measurements and predict their body size from 2D images taken with commercially available smartphones, we used machine vision and learning techniques. To estimate the body measurements of a human, the model 1) recognizes the human body in the images, 2) extracts the features of the body from the image, 3) determines the focal points of the human body, and 4) calculates the body measurements by computing the difference between the focal points. To predict the correct clothing size of a particular person, the models use a Support Vector Machine trained on some body measurements and body sizes of customers.

6.Conclusion This study proposes an approach that aims to improve and facilitate the online shopping experience by estimating a person's body measurements from 2D images by photographing the body with a smartphone camera. The experiment was conducted with a sample of volunteers who were photographed, manually measured, and asked to provide their actual clothing size to compare the result to the size predicted by the model. For the study, one of the pre-trained computer vision algorithms was used to detect the human body in images. The detectors are designed to detect three parts of the human body: one detector to detect the upper body, another detector to detect the lower body, and the last detector to detect the whole body. After detecting the main body parts, we extract features by segmenting each image into 40 parts and designating two points as focal points of each body part to estimate shoulder width, chest circumference, waist circumference, and hip circumference. Then, we used different machine learning models trained on a dataset of measurements to predict the clothing size as a function of the estimated measurements. Each model was trained to predict the size of a garment (i.e., to predict the size of a top, a bottom, or an entire garment). The results show that most of the predicted sizes have some deviations from the actual measurements of the participants. In the future,will work on improving the page image recognition result and use it to improve the size prediction, focusing on reducing the error percentage.

#### Estimating Human Body Dimensions Using RBF Artificial Neural Networks Technology and Its Application in Activewear Pattern Making:

#### Publication Year: 2019

**Author:** Zhujun Wang 1,2,3,4, Jianping Wang 1,5,\*, Yingmei Xing 2,3,4, Yalan Yang 1 and Kaixuan Liu 6

**Journal Name:** Appl. Sci. 2019

**Summary:** Abstract: Nowadays, the popularity of the Internet is constantly increasing. Intelligent prediction of human body measurements would be beneficial to companies in the garment industry to improve the precision and efficiency of pattern making. This study presents a new prediction model for estimating body measurements for pattern making in the apparel industry based on artificial neural networks (ANNs) with radial basis function (RBF). The model presented in this study was trained and tested using the anthropometric data of 200 adult males aged between 20 and 48 years. The detailed body measurements related to pattern formation could be obtained by inputting four easy-to-measure key measurements into the model RBF ANN. From the simulation results, the three-layer model with 4, 72, and 8 neurons in the input, hidden, and output layers had the highest accuracy when the scattering parameter σ and momentum factor α were set to 0.012 and 1, respectively, after being trained with a data set of 180 samples. Moreover, the prediction performance of the model presented in this study RBF ANN was better than that of the other two models according to the mean square error compared with a classical linear regression model and the backpropagation model (BP) ANN. Therefore, the presented prediction model is suitable for apparel pattern design, especially for close-fitting apparel patterns such as activewear. The estimation accuracy of the proposed model could be further improved if it were trained with more suitable datasets in the future.

Introduction In recent years, the requirements for individualized garments have greatly increased, including clothing styles, colors and fabrics. However, excellent fit is indispensable, as it is considered a crucial factor for the comfort of garments. In today's garment industry, pattern making is an important process to produce well-fitting garments. Patterns, also called paper patterns, are templates made of paper or cardboard that are used to draw the parts of the garment onto the fabric before they are cut out. Pattern making, sometimes called garment structure design, pattern design, pattern drafting, or pattern cutting, is a complicated technique that involves a ----- wide range of knowledge (e.g., esthetics, mathematics, ergonomics). The main problem in making well-fitting patterns is the design of garment sizes or dimensions, which are highly dependent on the knowledge and experience of pattern makers. Generally, garment sizes are designed and fitted by pattern makers based on human body measurements. Tailored garments require more accurate body measurements. Therefore, anthropometric measurement is an essential requirement for pattern making. With the increasing popularity of the Internet, shopping over the Internet has become an integral part of people's lifestyles. It is challenging for garment manufacturers to offer customized garments that are tailored to the exact body size and shape of a particular customer, as it is difficult to directly and accurately obtain a person's body measurements over the Internet. Normally, body measurements can be taken manually or automatically. With the advantages of intuitive and convenient tools, manual measurement with tape measures has been the traditional method to measure the human body for years. However, since this method is highly dependent on the experience and judgment of the people taking the measurements, the accuracy of the data obtained is unreliable, which can easily lead to the problem of poor fit of clothing. In addition, the method is time consuming. Compared with manual anthropometric measurement, 3D scanning technology for human body has greatly improved the efficiency and accuracy. In the past decade, various types of 3D body scanners have been launched and applied in the garment industry, such as laser scanning, patterned light projection, stereophotogrammetry, millimeter waves, and infrared waves [1,2]. Thanks to the development of charge-coupled devices (CCD), 3D body scanners have the advantages of high resolution (1-8 mm) and speed (0.2-3 s), which enable accurate and cost-effective acquisition of whole-body data [1,2]. Some other disadvantages of the devices, including bulky design, high price, and large storage capacity for 3D images, have also influenced their further application in the garment industry, especially in medium and small garment companies [2,3]. In the context of Internet sales, it is obviously unrealistic to use the two anthropometric measurement methods mentioned above [4]. In addition, clothing companies should meet consumers' individual needs as soon as possible. Therefore, the key measurements (e.g., height, chest circumference, waist circumference) that are easy to measure are measured physically, while the other detailed measurements are calculated by inputting the key measurements into empirical formulas based on linear regression models (LR). For example, sleeve length could be calculated by entering body height into a LR model. Due to their simplicity, LR models are widely used in industrial product design, e.g., clothing, tools, furniture, and workstations [5-9]. However, these models are not accurate enough [9]. Therefore, it is necessary to develop an approach to obtain body dimensions for garment manufacturing faster and more accurately than current methods. With the rapid development of artificial intelligence (AI), the prediction of human body measurements by AI instead of physical measurements has attracted more and more attention in the apparel industry. Due to the advantages of artificial neural networks (ANNs), such as good nonlinear approximation capabilities and adaptive and self-organizing abilities, and as one of the most popular machine learning approaches, ANN technology has been applied to many fields, including nanofluid viscosity prediction, human behavior prediction, pattern recognition, and adaptive control [10-14]. In the field of clothing pattern production, Chan et al [15] presented an artificial neural network model for predicting the pattern parameters of men's shirts in 2003. However, the inputs of the proposed model consisted of 58 body measurements, which were quite complicated and could not be easily acquired simultaneously. In 2014, a study by Zheng Liu et al [16] proposed a non-linear model to predict the detailed body measurements using feature parameters extracted by principal component analysis. But the parameters used as inputs for the proposed model were difficult to calculate in the study. Another study by Kaixuan Liu et al [17] in 2017 developed a neural network model with backpropagation (BP) to predict lower body size, which is used for designing pants patterns. However, the application of other neural network models was not mentioned in this study. Moreover, little attention has been paid to the application of the radial basis function (RBF) ANN in the creation of clothing patterns. Therefore, the aim of this paper is to present a new ANN model based on the radial basis function to improve the accuracy of body dimension estimation in garment manufacturing. Moreover, the proposed model can be used by pattern makers with little expertise and experience by inputting the learning data based on the knowledge of experienced pattern makers.

2. Methodology The proposed approach for estimating body measurements for making clothing samples is described in Figure 1. The detailed implementation process is as follows. First, anthropometric data were collected from a group of 200 men to create the database of human body measurements after preprocessing the data. In the next step, the data in the database were divided into two groups: the key dimensions and the detailed dimensions that are difficult to measure. Then, the predictive models ANN were developed with different mathematical algorithms that used the key dimensions as input variables and the detail dimensions needed to create clothing patterns as output variables. The created models were then trained and tested with the training dataset and the test dataset selected from the database. Then, the prediction performance of the proposed RBF ANN model was compared with the BP ANN model and the linear regression model.

As a popular garment, a tight-fitting activity garment without comfort additive was selected as a study sample to further test the performance of the presented model. The main body measurements of a new subject were entered into the model. Then, the body measurements needed to create the activewear sample were estimated by the model. Since the selected activewear did not include a comfort allowance, the output body measurements could be used as the pattern. The activewear pattern was finally created based on the estimated body measurements.

#### SmartFit: Smartphone Application for Garment Fit Detection Publication Year: 2021

**Author:** Kamrul H. Foysal 1 , Hyo Jung Chang 2 , Francine Bruess 2 and Jo Woon Chong 1,

**Journal Name:** Electronics 2021,

**Summary:** Abstract: The apparel e-commerce industry is growing day by day. Recently, consumers are especially interested in an easy and time-saving way to buy clothes online. In addition, the COVID -19 pandemic has increased the need for an effective and convenient online shopping solution for consumers. However, online shopping, especially online shopping for apparel, poses several challenges for consumers. These include issues of size, fit, returns and cost. Especially the issue of fit is one of the main factors that cause hesitation and disadvantage when buying clothes online. The traditional method of determining the fit of clothes based on body shape relies on manual body measurements. Since there is not yet a convenient and easy-to-use body shape detection method, we propose an interactive smartphone application, "SmartFit", that recommends the optimal fitting clothes to consumers by detecting their body shape. This optimal recommendation is created using image processing and machine learning based solely on smartphone images. Our preliminary evaluation of the developed model shows 87.50% accuracy in body shape recognition, which is a promising solution to the problem of fit recognition in the digital apparel market.

Introduction Recently, society's demand for buying clothes online has increased with the technological progress. As a result, the apparel e-commerce industry is booming more than ever [1]. Therefore, the online presence of apparel companies is crucial to the success of the business. This is especially true during the COVID -19 pandemic, while online apparel shopping for consumers is rapidly and gradually increasing [2]. E-commerce sales of apparel are $110.6 billion in 2020 and are expected to increase to $153.6 billion by 2024 [3]. However, one of the main obstacles to increasing sales in this industry is the large number of returns. In addition, the industry has problems with marketing, shipping, and processing costs [4]. A study has shown that about 30% of garments purchased online are returned because of poor fit [5]. Currently existing online shopping websites do not provide a way to measure body size and shape. Tech startups such as True Fit (www.truefit.com) and Fits.me (www.fits.me) have developed virtual fitting rooms where virtual mannequins mimic customers' body measurements and display how a particular garment, i.e., a garment, might look when worn by the customer [6]. However, these types of solutions fail in practical sizing at the industrial level due to technical hurdles such as an overly complicated process, inefficient technology, low user customization, and costly expenses to implement the technology [7]. Even advanced technologies such as Sizer (www.sizer.me) and MTailor (www.mtailor.com) do not always provide actual fit information and size measurements and cannot completely solve the problem of poor fit [8].

Detecting the fit of online clothing is a burning issue because there is no technology that can detect consumers' body shapes. The fit of garments largely depends on the individual body shapes of consumers [9]. Therefore, there are huge profit losses due to poor fit and related return problems. In fact, online purchases result in returns 50% of the time, costing both the consumer and the retailer time and money. Returns can cost a total of about $10-15 per garment, which is a huge burden for consumers and businesses [10]. Since most garments are universally made for traditional body shapes and sizes, there is a need to address fit issues in individual ways, starting with garment manufacturers and ending with retailers and consumers. In addition to custom fitting in clothing stores, clothing manufacturers are moving to produce garments based on anthropometric data [11,12]. The four most common traditional body shapes for women [13], namely the inverted triangle, pear, hourglass, and rectangle [14], are shown in Figure 1. Body shape plays an important role in understanding the optimal fit of garments for individual consumers [15]. However, current technologies cannot provide information about body shape in relation to garment fit [10].

In this work, a novel body shape recognition algorithm for garment fit was proposed that extracts consecutive feature points using accelerated robust features (SURF) [16], recognizes visual words from the extracted feature points using the bag-of-features model [17], and classifies body shapes using a machine learning technique such as k-nearest neighbor (k-NN) [18]. In addition, a body shape recognition algorithm based on a Convolutional Neural Network (CNN) [19] was proposed. Specifically, a unique combination of image processing and machine learning was proposed based on body shape recognition using 2D smartphone images. In particular, feature recognition ( SURF ), bag-of-features model, and machine learning technique have not been previously proposed for discriminating body shapes for garment fit. The conventional body shape recognition method is limited to a physical measurement-based approach [20] or a multiple photo-based 3D reconstruction approach [15], which are generally inaccurate and impractical due to expensive computations [21]. Therefore, in this work,we propose a unique combination of image processing and machine learning based on body shape recognition using 2D smartphone images. Figure 2 describes the general approach of the proposed method. -----

Conclusion In this article, we propose an effective body shape detection method using a smartphone that solves fit problems, increases consumer convenience, and improves the online clothing shopping experience for consumers. The proposed method was shown to provide an average accuracy of 87.50% in classifying four different body shapes: Inverted Triangle, Pear, Hourglass and Rectangle. The proposed methods are expected to increase the feasibility of smartphone-based online clothing shopping by suggesting garments with optimal fit in a more accurate and personalized manner [52], which ultimately benefits online clothing retailers by increasing their sales and reducing returns due to poor fit. The results of this study demonstrate the potential contribution to the textile and apparel industry by enabling non-contact body shape recognition with high accuracy. Further development of the proposed method will produce a model that enables the recognition of other body types by incorporating additional data on other existing body shapes. The next step in this research is to analyze the shape and fit of the garment itself to match it with the body shape recognition data so that information about the optimal fit of the garment can be provided to the consumer.

#### SizeNet: Weakly Supervised Learning of Visual Size and Fit in Fashion Images:

#### Publication Year: 2019

**Author:** Nour Karessli Romain Guigoures Reza Shirvany ` Zalando SE - Berlin, Germany

**Journal Name:** ---

**Summary:** Abstract Finding clothes that fit is a hot topic in the e-commerce fashion industry. Most approaches that address this problem are based on statistical methods that rely on historical data of purchased and returned items. Such approaches suffer from the cold start problem with the thousands of items that appear on shopping platforms every day for which no previous purchases are available. Therefore, we propose to use visual data to determine the size and fit of fashion items. We present SizeNet, a weakly supervised teacher-student training method that leverages the power of statistical models combined with the rich visual information of item images to learn visual cues of size and fit characteristics that can solve the challenging cold start problem. Detailed experiments are conducted on thousands of garments, including dresses, trousers, knitwear, tops, etc. from hundreds of different brands.

1. Introduction The apparel industry has been an important contributor to the economy of many countries. In particular, fashion e-commerce has become a major player in recent years, providing competitive and customer-oriented products and services. Recent studies have shown that finding the right size and fit are among the most important factors influencing customers' purchase decision process and their satisfaction with e-commerce apparel platforms [1]. Online shopping requires customers to buy clothes without trying them on. This naturally delays the sensory feedback stage about the item's fit through touch and visual cues, leading to uncertainty in the purchase process. As a result, many consumers are hesitant to engage in the buying process, especially with new items and brands they are unfamiliar with. To make matters worse, there are significant size differences in fashion items, including footwear and apparel, largely due to: 1. A rough definition of size systems for many categories (e.g., small, medium, large for apparel); 2. Different indications for the same size depending on the brand; 3. Different ways of converting one local size system to another, e.g., in Europe, apparel sizes are not standardized and brands do not always use the same conversion logic from one country to another. One way to get around the confusion caused by these differences is to provide customers with size charts that convert aggregated body measurements into the item size system. However, this requires customers to record their body measurements. Interestingly, even when customers are provided with accurate body measurements with the help of individual instructions and expert explanations, the size charts almost always have a high variance, which can be as much as one inch for a single size. These differences either stem from different data sets used to create the size tables (e.g., German vs. British population) or are due to vanity sizes. The latter is when a brand intentionally creates size inconsistencies to satisfy a specific customer group based on age, athleticism, etc., which have a large impact on the body measurements reported in the size tables [2, 3, 4]. The combination of the above factors presents customers with the difficult problem of determining the correct size and fit for purchase. In recent years, there has been a lot of interest in developing recommendation systems for fashion e-commerce, focusing on modeling customer preferences based on their past interactions, tastes, and inclinations [5, 6, 7]. Other work has focused on image classification [8, 9], fashion product tagging and discovery [10, 11], algorithmic outfit generation and style extraction [12], and visual search, which focuses on the problem of matching studio and street images to e-commerce items [13, 14]. In this context, very little research has been done to understand how fashion items behave from the perspective of size and fit [15, 16, 17, 18], with the main goal of providing size advice to customers, mainly by leveraging similarities using item sales and returns data, as described in Section 2. Returns have various reasons, such as "the item does not like, the item is damaged, the size is not right, etc.". We propose a weakly supervised [19] teacher-student approach [20, 21, 22] that first uses item sales and size-related returns to statistically model whether an item suffers from size problems or, conversely, has normal size and fit behavior. In this context, you do not have access to expert-labeled data on item size and fit and therefore rely only on weakly annotated data from the returns process. Therefore, we use a teacher-student approach with curriculum learning [23], where the statistical model acts as the teacher and a CNN-based model, called SizeNet, acts as the student, with the goal of learning indicators of size problems from fashion images without direct access to privileged sales and returns data. Our work makes a threefold contribution: 1. Demonstrate for the first time, to our knowledge, the great value of fashion images in inferring size characteristics of fashion apparel; 2. At the same time, the approach is novel in that it uses image data to effectively address the cold start problem known in the literature to be very difficult; 3. Propose a statistical teacher model that leverages subjective and imprecise audience feedback (which is highly influenced by personal perceptions of item size) to generate confidence-weighted weak annotations on a large scale. This allows us to control for the extent to which the weak annotations affect the quality of the final model. Thus, we show that not using this approach, i.e., treating the weak annotations uniformly, greatly degrades the quality of the learned model.

#### Clothes Size Prediction from Dressed-human Silhouettes: Publication Year: 2017

**Author:** Dan Song1 , Ruofeng Tong1? , Jian Chang2 , Tongtong Wang1 , Jiang Du1 , Min Tang1 and Jian Jun Zhang2

**Journal Name:** ------

**Summary:** Abstract. In this study, propose an effective and efficient way to automatically predict clothing sizes for users who buy clothes online. We take human body size and the silhouette of the clothed person in front and side views as input and estimate 3D body sizes using a data-driven method. We assume 20 body sizes that are closely related to the dress size, and use these 3D body sizes to estimate the dress size by searching the corresponding size chart. Previous image-based methods need to calibrate the camera to estimate 3D information from 2D images because the same person has different silhouettes (e.g., height and shape) when the camera configuration (intrinsic and extrinsic parameters) is different. The method does not require camera calibration, which is much more convenient. So, set up our virtual camera and train the relationship between human size and silhouette size under this camera configuration. After estimating the size of the silhouette, regress the positions of the 2D body landmarks. Thus, define the 2D body sizes as the distances between the corresponding 2D body features. Finally, we learn the relationship between the 2D body sizes and the 3D body sizes. The training examples for each regression process come from a database of clothed 3D bodies created in previous work. evaluate the entire procedure and each process of our framework and compare the performance with different regression models. The total time required to predict clothing size is less than 0.1 seconds, and the average error in estimating body size is 0.824 cm, which accommodates customers' tolerance for online clothing purchases.

1 Introduction The right dress size plays a crucial role in online dress shopping success. However, it is not easy for customers to find the right size when buying clothes online. On the one hand, customers must have professional skills to measure themselves with a special tool such as a tape measure in a relatively private space. On the other hand, customers need to check the size charts of the items because the size standards are different in different countries and different clothing brands. In a word, the techniques for automatic size suggestions are in demand. We assume 20 body sizes (Figure 2 (b)) related to clothing size, including the length and circumference information of 3D bodies. For these 20 body sizes, the dress size can be obtained by searching the corresponding size table. Therefore, we focus on automatic estimation of 3D body sizes. Images contain valuable information and are quite easy to obtain, . In general, there are two ways to predict body sizes from images. One way is to measure 3D body shapes reconstructed from images, while another way is to use body features extracted from images. Image-based reconstruction of the human body has attracted many researchers. Some researchers learn the relationship between 2D positions in the image and the corresponding depth information using a database of 3D bodies. Other researchers train a parametric human body model with a body database and use images as constraints to deform a template mesh. Some methods integrate both body reconstruction ideas presented above. They first estimate rough bodies based on silhouettes using machine learning methods. Then the bodies are geometrically refined to match the body contours of the silhouettes. Body sizes are determined by measuring the 3D bodies reconstructed using these methods. In some works, landmarks are first determined from the body contours and then the body size is predicted based on the landmarks. Several methods have been presented to detect the position of landmarks, but they have their limitations. Manual labeling requires a professional sense of landmark location and costs time and energy. In corner point detection, the landmarks must be corner points and are very sensitive to the quality of the silhouettes. Some researchers extract landmarks using Iterative Closest Point (ICP) for 2D curves, which is very time consuming. Song et al [1] use a 3D regression method for landmarks, which is very efficient and insensitive to body contours, to exploit global information. However, they require camera calibration to regress 3D landmarks from 2D silhouettes. Usually, machine learning methods are used to predict 3D body sizes from body features. Similar to the regression method for 3D landmarks [1],we regress 2D landmarks from silhouettes. However, we do not calibrate the camera, but require the height as input, which is much easier to obtain. The camera configuration (intrinsic and extrinsic parameters) affects the size and shape of the human silhouette. The shape of the silhouette is only slightly affected if the camera is relatively far from the center of the human body in orthogonal view. The size of the silhouette is estimated using the height with a fixed camera configuration. Then, regress the positions of the 2D landmarks from the scaled silhouettes. Unlike [1], we do not reconstruct the human body in 3D, but learn the body size using 2D landmarks. therefore introduce the framework and the main processes and evaluate them in . For the specific task of body size estimation, the advantages ofthe method can be summarized as follows: (a) automatic, (b) efficient (the total time required is less than 0.1 seconds), (c) effective (the average error is 0.824 centimeters), and (d) free of camera calibration. Previous methods can only meet some of the above advantages. Moreover, our method can be easily integrated into shopping websites for clothes.

FashionFit: Analysis of Mapping 3D Pose and Neural Body Fit for Custom Virtual Try-On

ABSTRACT Visual compatibility and virtual sensation are critical criteria for fashion analysis, but are missing from existing fashion designs and platforms. An explicit model is urgently needed to ensure visual compatibility by coloring fashion images and virtual try-on. With rapid advances in computer vision and the increasing creation of customer experiences, there is great potential for retailers and customers. Available public datasets are well suited to generate outfits from Generative Adversarial Networks (GANs), but user-generated outfits have low accuracy. This work is the first step in analyzing and experimenting with the fit of custom outfits and visualizing them for users, resulting in a great customer experience. The work analyzes the need for visualization of custom outfits for users in the large corpora of AI in Fashion. The authors propose a novel architecture to combine outfits provided by retailers and visualize them for users themselves using Neural Body Fit. This work sets the standard in disentangling custom generation of garments with GANs and virtual fitting on the user to ensure a virtual-photorealistic look and result and create a great customer experience using AI. Extensive experimentation shows that the outfits generated by GANs have high accuracy, but not in the customized execution. This experiment creates new state-of-the-art results by recording the user's pose to calculate the lengths of each body part (hand, leg, etc.), as well as segmentation + NBF for accurate outfit customization. This work is different from all other competitors in terms of virtual fitting approach to create a new customer experience.

A Review of Body Measurement Using 3D Scanning

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9419003

https://virtualhumans.mpi-inf.mpg.de/papers/sattar2019WACV/sattar2019WACV.pdf

https://openaccess.thecvf.com/content\_CVPR\_2020/papers/Hsiao\_ViBE\_Dressing\_for\_Diverse\_Body\_Shapes\_CVPR\_2020\_paper.pdf

#### Summary of Literature Survey:

Here, explored different machine learning approaches for predicting fit and body measurements. All approaches have their own advantages and disadvantages. Fit recommendation with a predictive model is the most popular algorithm for prediction, but this method has some challenges, such as the need for a large number of training data, partitioning of data, lack of adequate and accurate training time, and dependence on previous information for prediction. A prediction model with an appropriate regression approach can be used to overcome these problems. Machine learning can provide highly accurate predictions using standard tools and outperforms all standard prediction methods.

**Chapter 3. Existing Work and Proposed Work**

### Overview of Existing Work

Personalized sizing and fit recommendations are critical to any fashion e-commerce platform. Predicting the right fit promotes customer satisfaction and benefits the business by reducing the cost of size-related returns. Traditional collaborative filtering algorithms attempt to model customer preferences based on their previous orders. A typical challenge for such methods is that the number of orders for customer items is very small. To alleviate this problem, a Deep Learning-based collaborative method for personalized size and fit recommendations was developed. The method can ingest arbitrary customer and item data and can model multiple people or intentions behind a single account. The method optimizes a global set of parameters to learn population-level abstractions of relevant size and fit information from observed customer-item interactions. It also uses customer- and item-specific embedding variables to learn their properties. Together with the learned embedding variables, the method maps additional customer and item attributes in a latent space to derive personalized recommendations. Applying the method to two publicly available datasets shows improvement over published state-of-the-art results. On two proprietary datasets, one containing fit feedback from fashion experts and the other purchases from customers, the method outperforms comparable methods, including a recent Bayesian approach to size recommendations.

The topic of understanding sizing problems and predicting size and fit at a personalized level has gained momentum in the research community. In what follows, we outline some recent developments on this topic and draw parallels between the work and closely related methods of collaborative filtering: the method of matching customer images to existing 3D body scans that are matched to items to generate fit ratings. The method presented in [1] proposes to apply a skip-gram based word2vec model [17] to purchase history data to learn latent representations of items. The approach then builds a customer representation by summarizing the learned representations of items purchased by customers. A gradient-boosted classifier is then trained on latent representations of customers and items to predict fit, thus a hierarchical Bayesian approach for personalized size recommendations. Given pairs of customers and items, the method models the joint conditional probability of the sizes ordered by the customers along with the outcomes observed in the training data (i.e., retained vs. size-based returns). To provide personalized size recommendations, the method uses the conditional probability of size given a customer and an item with a "keep" outcome. The method uses approximate probabilistic inference for parameter optimization and testing.

In the following , we derive the "true" sizes of customers and items from the purchase and return data using a latent factor model. The derived size features are fed into a standard classification system to perform ordinal prediction of fit (i.e., 'Small', 'Fit', 'Large'). The method also performs hierarchical clustering for individual customer data to handle multiple customers behind an account. The method relies on approximate probabilistic inference (mean-field variational approximation with Polya-Gamma augmentation) to estimate the posterior distribution over customer and item sizes. A conceptually similar work that models the size recommendation problem as an adjustment prediction problem. In a two-step procedure, the method first learns to embed customers and items in a latent space with the same dimensionality. Once the embeddings are obtained using an ordinal regression procedure, they are used in the next step to learn representations for each class by applying prototyping and metric learning techniques. The public datasets we use for the benchmarking approach. Most of the above works do not take an end-to-end approach to the posed task, while some are limited in terms of scalability (e.g., due to their probabilistic nature) or capacity (e.g., due to predefined interactions, linearity assumptions, the ability to handle cold starts, or to model multiple users/intentions behind an identity). Our work, on the other hand, presents a scalable, end-to-end Deep Learning approach to size and fit recommendations. The two-path neural network architecture used in this work (Figure 1) flexibly processes both categorical and continuous customer and item features, and learns (potentially non-linear) customer-item interactions from the data. Our model architecture is quite general in the context of collaborative filtering. For example, it is closely related to the Deep Structured Semantic Model (DSSM) [10] and Neural Collaborative Filtering (NCF) [9]. The DSSM was developed for web search and uses independent neural network layers to embed customers and articles in a latent space. It then uses a predefined interaction between the latent embeddings to predict its target. NCF uses an architecture inspired by Neural Tensor Networks [24] to learn input embeddings or features for (point-coded) customers and items. The architecture includes both shallow (GMF) and deep (MLP) feedforward paths to model both linear and nonlinear interactions between pairs of customers and items. A notable difference between our architecture and DSSM or NCF is that our architecture uses skip connections [8] between layers. Our proposed approach can be viewed as a generalization of logistic matrix factorization [12], which is a linear model of customer-item interactions. Apart from the interaction data, the method does not consider any other customer or item information for generating personalized recommendations.

Clothing is important for the wearer's identity and nonverbal communication. The comfort of clothing depends on many social and physical factors. The fit of clothing affects comfort, and a well-fitting garment improves both appearance and self-confidence. Providing well-fitting clothing requires many tools and resources. These include body measurements, pattern technology and garment fit simulation. These are used in the product development process to improve and evaluate garment fit and to assist in size selection. The importance of clothing Clothing is not only used to cover and protect our bodies. Today, clothing also serves to identify us (Alexander, Connell & Presley 2005). Clothing conveys personality, points of view, group membership, etc. (Feather & Jenkins 1993; Jacobson 1994; Ryan 1966) and is crucial to a person's first impression (Molloy 1988; Thorén 1992). According to Ashdown and O'Connell (2006), well-fitting clothing improves appearance and increases self-confidence. Although many authors emphasize that clothing is a powerful communication tool, Feinberg (1992) believes that not all clothing sends clear and complete signals about the wearer's self-identity because interpretation depends on the observer (Ryan 1966). Because clothing is so important to self-identity and nonverbal communication, it is of great interest for everyone to find appropriate and comfortable clothing for each occasion. Clothing Comfortable clothing is different from other products on the market. Since it is worn close to the body (Ashdown 2014), the comfort of the product is important. The definition of clothing comfort is difficult to quantify (Slater 1986) and even difficult to define (Li 2001). The factors that influence clothing comfort are interactive, and individual evaluations vary, making the topic difficult to study (Goldman 2005). The literature describes clothing comfort in different ways; they overlap and use different vocabulary, making comparison difficult. Nevertheless, Figure 1 provides an overview of clothing comfort. Sontag (1985) begins by placing comfort in a context in which the environment, clothing factors, and people interact. The environment includes attributes such as climate, people, and social norms. Clothing factors include all the physical attributes that influence comfort. For example, people relate to personal values, age, ethnicity, gender, height, weight, and physiological aspects (Sontag 1985). All of these aspects are considered when evaluating comfort. Individuals consciously or unconsciously evaluate both physiological and psychological comfort. Physiological comfort includes, for example, the state of body temperature, lung function, tactile sensation, blood pressure, and visual and auditory stimuli. Psychological comfort includes the characteristics of the environment in relation to people and social norms. (Slater 1986) The social and physical factors influence the evaluation of clothing comfort. The social factors that influence clothing comfort relate to personal experiences and how others react to the individual. Psychological comfort may be met when the individual feels that he or she is wearing the right clothing for a particular occasion (Sontag 1985). The literature describes four main physical factors that affect physiological and psychological comfort: thermal, sensory, ergonomic, and esthetic (Das & Alagirusamy 2010; Roy Choudhury, Majumdar & Datta 2011).

<https://www.diva-portal.org/smash/get/diva2:1216128/FULLTEXT01.pdf>

Online shopping platforms offer consumers the opportunity to store from the comfort of their own homes. Fashion, and especially apparel, is the fastest growing category in online shopping [1]. There are several factors such as trust, logic, fit, and emotion that influence the user's choice when purchasing clothing through online portals [2]. It has been shown that the fit of the clothing is the most important element for consumers to determine their overall satisfaction with the clothing [3]. When the size and fit are not right, consumers often return the purchased garments [4]. To facilitate online shopping, most online e-commerce platforms offer free returns on all their products. Online e-commerce platforms typically experience nearly twice the return rates of traditional offline fashion stores [5]. During the return process.

The reason for the return by the customer. Using this data, we can attribute a large percentage of returns to size and fit errors. Returned products cause significant operational costs on the platform and block inventory for sale. In addition, this creates a poor experience for customers and lowers their confidence in future purchases on the platform. Compared to traditional offline business, in online business there is no physical product to check and try. The consumer's purchase decision is based solely on images, descriptions and size charts attached to the product. Size charts require customers to remember their body measurements and compare them with the product measurements. Moreover, there is no standardization in sizing in the fashion industry [6] and the attributes associated with fashion products are very subjective. Each clothing category tends to have similar size specifications - S, M, L, XL, etc. across all brands, but representing different body measurements. And even within the same brand, different product lines and different fits (slim, regular, etc.) make sizing a difficult process. In addition, the actual product measurements and the size charts provided by the brands differ greatly. During quality control, we measured samples for each product and found that 30% of the garment inventory had a deviation of ± 1 inch compared to the size chart. All these factors make selecting the right size from a size chart very tedious. This often leads to a negative experience for customers and hinders adoption by new customers. There is extensive research on personalized product recommendations for users in fashion e-commerce. In apparel and textile research, several methods have been proposed for determining fit and size preference based on 3D modeling of body shapes [11; 12; 13]. These methods rely heavily on deriving body shapes from databases of manually curated body shape metrics [14] or extracting body shapes from images [15]. There have also been attempts to model user size preferences in the industry, e.g., by True Fit and Fit Analytics. However, these approaches require users to explicitly provide their body measurements in surveys or questionnaires. https://heindaanen.nl/images/techupdate8.pdf

The project 'Passende mode via Internet' has the im (fitting fashion using the internet) to reduce the number of unwanted returns of clothing. The Netherlands is currently the record holder for the number of returned goods, with fashion leading the way (https://www.eshopperbarometer.dpdgroup.com/chart.php). Improper fit is a major reason for returning merchandise. The project, method for achieving the goal of reducing fashion returns is to identify the customer's preferences and body measurements and translate that information into the selection or production of a best-fitting garment or other fashion item. Five solutions have been identified and will be discussed sequentially in this report: 1) smart questions that the consumer answers over the Internet so that an image of the consumer's body measurements is retrieved (and ultimately the correct size garment can be delivered) 2) the use of a smartphone or tablet to derive body measurements 3) the use of 3D body scanners 4) converters or plug-ins that ask for an existing garment that fits well and convert that information into body measurements. Finally, the possibilities of 3D printing garments are briefly discussed.

https://www.researchgate.net/publication/325172567\_Garment\_Fit\_Evaluation\_Using\_Machine\_Learning\_Technology

Nowadays, electronic purchasing of garments has gained importance worldwide [1] . However, an important technical obstacle is that garments offered online cannot be physically tested for their effect on a particular consumer [2] . Therefore, virtual try-on technology has been developed to evaluate the quality of garments [2,3] and has been widely used in the apparel industry over the past decade. A number of virtual try-on programs such as Clo 3D, Lectra 3D Prototype, OptiTex, and V-Stitcher 3D are available in the market to evaluate the quality of garments [4]. These 3D virtual try-on systems follow similar principles, i.e., they show the static and dynamic behavior of a virtual garment based on identified human morphological and fabric properties and their interactions. The development of this behavior requires the use of complex mechanical and geometric modeling and simulation techniques, such as simulation with ﬁnite elements [5].

https://heindaanen.nl/images/techupdate8.pdf

Basic body assessment

Basic Body Assessment Some body measurements are well estimated by most subjects, such as body weight and stature. Age and gender are also known and related to body measurements. Using these values alone can provide a rough estimate of body shape and some other body measurements such as inside leg length. The addition of collar width, which is known to most men, leads to a reasonable estimate of body measurements that are important for shirt design. Bivolino (www.bivolino.com) applies this technique to men's shirts. The estimated upper body measurements are transferred into a customized clothing pattern. It is well known that estimated weight and stature do not always match measured values. Heavy people tend to underestimate their body weight, and short and elderly people tend to overestimate their stature. These systematic over- and underestimates are published and can be used to correct estimated values (Krul, Daanen, & Choi, 2011).

Self‐assessed body dimensions

Self‐assessed body dimensions Another option is to ask people to assess the length or circumference of certain body parts without measuring them. For instance a small slider on the screen can be used. The idea behind this method is that people continuously compare themselves to peers and experience in shops how they compare to body dimensions of other people. This implicit reference frame can be quantified by asking smart questions or use tools like sliders. There seems to be a fairly good relation between self‐assessed body dimensions and measured body dimensions, e.g. for arm length (H.A.M. Daanen & Byvoet, 2011). At the bivolino website (www.bivolino.com) customers can indicate if they experienced body dimensions or shapes deviating from normal, and if so they can indicate which. It is better to ask for clear body dimensions than to ask people to compare themselves to shapes like rectangles or triangles. Also asking if someone is pear shaped or shaped as a hourglass (e.g. http://shopyourshape.com/body‐shapes/) does not yield useful info. Sizebuddy (www.sizebuddy.com) is developing a passport in which essential body dimensions are stored. When the internet is searched for garments, only the applicable sizes are shown, saving time and money. It may help if the body is visualized on the screen and if the body dimensions can be adapted by the user. Stylewhile (www.stylewhile.com) is an example, as well as virtual manikin (www.). In Amsterdam, the company MimicMe was active for a while, but this is no longer the case. Several companies ask smart questions on other topics than body dimensions in order to propose garments that are in line with customer interest. https://www.silksage.com/ for instance wants to know if you are always in a hurry or not, if you are creative or not and uses this information in the proposal of fitting garments.

2.4 Self‐measuring of body dimensions

Asking a client to measure themselves may produce inconsistent or incorrect results. For example, the tape is old and stretched, the person measures on the wrong part of the body, or reads the length of the tape on the wrong side. Therefore, good instruction and some training is essential. Traditional methods of body measurement are described in the standards ISO. ISO 8559 is dedicated to body measurements for garment design (http://www.iso.org/iso/catalogue\_detail.htm?csnumber=15821). Reproducible measurement of the human body requires training. ISAK (http://www.isakonline.com/courses) provides this training. It is good to know that it is not possible to train all customers. The Besized application (www.besized.com) is a personal and brand-specific tool for sizing children's clothing that benefits both consumers and web stores. The customer enters the child's date of birth, stature and weight. Based on this information, clothing sizes are displayed. The idea is that the shopping experience for the customer improves as Besized helps them choose the right size. In this way, the hurdle of size uncertainty, which in most cases prevents consumers from buying, is overcome. The idea is that Besized significantly reduces the cost of returns and improves the image of a website. Besized uses scientific data (TNO / Netherlands Organisation for Applied Scientific Research). Besized has a large database with representative samples of children's measurements in the Netherlands and specific size charts of different brands. Using these data sets in a specific model result in a size recommendation for children's clothing.

Smartphone or Tablet

1. Several companies offer apps on a smartphone or tablet that make photos from which body dimensions can be deducted. IBV (Institute Biomechanics of Valencia) produced an app called Kidsize (www.kidsizesoluition.com) that asks for a frontal and sagittal picture as input, as well as information on age and stature. Stature is used for scaling. The app is developed in a EU project. The profiles are linked to a 3D scan database of 800 children. The tool is thoroughly validated, and a link to garment size is made using 1100 fit tests with children. The fit assessment is done by independent experts and parents. The size recommendations are good; it is one of the best tools available. They are expanding the app to adults. The results are accepted for publication in the Elsevier journal ‘Data in brief’. Another company that follows this line and employs the same technology is QuantaCorp (www.quantacorp.com). The first draft was evaluated by Vonk and Daanen (Vonk & Daanen, 2015). A new version is available that may be evaluated in this project. In a more recent evaluation, the QuantaCorp system was used to predict the best fitting size of a Jobe Unify vest https://www.jobesports.com/jobe‐unify‐vest‐men‐black‐244917102/ that was fitted in 22 males and 39 females. The QuantaCorp size prediction was close to the expert for males and slightly bigger for females. Submission of this study to a scientific journal is planned. Astrivis (www.astrivis.com) uses a smartphone for 3D scanning. It is nice for objects, but seems less suited for body (parts). Development is continuing. Netvirta (www.netvirta.com) uses a smartphone to scan (parts of) the body as well. Nettelo (www.nettelo.com) uses a smartphone to determine body dimensions and interact with the user.
2. 3D about me (www.3dabout.me) uses an app to scan the feet and supply the fitting size of the shoe. They have a database for major brands on the conversion of foot dimensions to shoe dimensions and currently also work for the Dutch military. Another company active in 3D shoe fitting is www.safesize.com. Metail (www.metail.com) offers virtual fitting on a copy of your body visualized on a smartphone or tablet. They offer the service to photograph the garments and then the customer can see it on her or his body. Yannis Douros, previously working on SizeUK, works for Metail. It is owned by the same Hongkong company that has major shares in Sizestream. A new concept is Fitbay (https://fitbay.com/). If a garment fits you perfectly you send this information to a friend with similar body dimensions using the fitbay app. The idea is that this way clusters of similar body shapes will develop that notify each other and thus looking for the right size becomes easier. http://mysizeid.com/ algorithms to capture a person’s measurements using their smartphone sensors, without the need to use a camera. You move your smartphone over the body and thus measure the body dimensions.
3. https://heindaanen.nl/images/techupdate8.pdf

https://towardsdatascience.com/would-this-clothing-fit-me-5c3792b7a83f

Body Mass Index and Body Satisfaction: Does Availability of Well-Fitting Clothes Matter?

Abstract

Relatively little is known about the factors that mediate the association between high body mass index (BMI) and lower body satisfaction. This is the first study to examine whether the presence of well-fitting clothing mediates this relationship. Eighty-five women aged 18-81 years were recorded, weighed, and measured at Time 1 using a 3D body scan, and the number of retailers carrying their sizes (determined by the body scan) was calculated. At time 2, they completed an online body satisfaction questionnaire. Body satisfaction at Time 2 was predicted by both BMI and the availability of well-fitting clothing at UK retailers at Time 1, with the two factors explaining 27% of the variance in body satisfaction. Clothing size availability partially mediated the relationship between BMI and body satisfaction. The results suggest that clothing retailers could help reduce body dissatisfaction by offering greater choice to consumers of all sizes.

## Evidence That BMI Predicts Body Satisfaction

BMI is a good predictor of body satisfaction in both men and women. Therefore , it is not surprising that higher BMI (weight to height) generally predicts body dissatisfaction in women (Bacevičienė et al., 2009; Burrowes, 2013; Streeter et al., 2012; Weinberger et al., 2016).

Body dissatisfaction has been defined as "negative thoughts and feelings a person has about their body" (Grogan, 2017, p. 4). Body dissatisfaction has negative implications for women: It is not only associated with lower self-esteem (O'Dea, 2012), but also with health risk behaviors

Weinberger et al. (2016) conducted a systematic review of 17 articles to compare differences in body dissatisfaction between adults who were "normal weight" and those who were "overweight." They found significantly higher body dissatisfaction in obese participants across all studies and a significant association between female gender and body dissatisfaction, such that women reported significantly higher body dissatisfaction even when their BMI was lower than that of men. These differences were found regardless of whether researchers used body satisfaction questionnaires (d = 0.89, 95% CI [0.63, 1.16], p <.001) or silhouette scales (d = 1.41, 95% CI [0.57, 2.25], p <.001).

Weinberger et al. (2016) pointed out that the questionnaire studies they examined were limited in that almost all respondents self-reported their height and weight information to calculate BMI. Although self-report is inexpensive compared to measuring weight and height, participants tend to underestimate their weight and overestimate their height when they self-report these measures, possibly due to the social desirability bias associated with social pressures to be tall and slim in Western societies (Elgar & Stewart, 2008). Another limitation is that previous researchers have tended to assess the relationship between BMI and body satisfaction cross-sectionally, potentially inflating the magnitude of the correlation between BMI and body satisfaction and potentially conflating the experience of being weighed and measured with concurrent body satisfaction. In the current study, we use direct measures of height and weight to calculate BMI and a prospective design to predict body satisfaction using baseline anthropometric measurements taken a few weeks earlier to avoid some of the limitations of previous work in this area. Thus, it can be predicted that higher BMI leads to greater body dissatisfaction.

## Clothes Sizing, Fit, and Body Satisfaction

Clothing stores usually offer only a limited selection of sizes. This lack of availability of clothing in larger sizes creates practical problems with clothing selection (Reardon & Grogan, 2011) and can also signal to women that their bodies are outside of acceptable size norms, causing them to become increasingly dissatisfied with their bodies.

There are many definitions of clothing fit, some from the perspective of clothing manufacturers and others from the perspective of consumers. Alexander et al. (2005) point out that the success of a garment's fit depends both on how the manufacturer interprets body measurements when designing the garment and on individual fit preferences, which may vary. Garment fit can vary according to fashion standards and the preferred fit of certain garments, which means that garment fit can be individual and subjective. For the purposes of this study, clothing fit is objectively defined as a wearer's body measurements (determined through body scanning) that are within +/- 3 cm of a retailer's sizing information (determined through online sizing charts) for bust, waist, and hip circumference (Gill, 2015).

Clothing fit can be used by women as a means to monitor their weight (Grogan, 2017), so it can have important implications for body dissatisfaction. Not fitting into their usual clothing size can signal to women that they have gained weight, which can lead to increased body dissatisfaction (Grogan et al., 2013). In interviews with 20 women aged 18 to 45 years on the topic of clothing and body image, Grogan et al. (2013) found that all of the women reported they paid close attention to their body size in relation to clothing fit, and that tight clothing resulted in lower body satisfaction, even among women who originally had high body satisfaction. Although the women in Grogan et al.'s (2013) study "knew that size measurements are unreliable," they still used clothing size as a marker for weight gain and were extremely unhappy when clothing was too tight and did not fit as expected (p. 387)..

## The Present Study

* This study was designed to investigate the relationship between BMI, body image, and the availability of clothing in an appropriate size. To avoid problems encountered in other studies, we (a) used a prospective study design in which we took body measurements at time 1 and measured body satisfaction at time 2, (b) drew a sample of women aged 18 years and older with no upper age limit to examine whether the observed associations were influenced by age, and (c) measured weight and height objectively with a scale and a Leicester height measure. To determine the number of retailers carrying clothing in appropriate sizes, the women's bodies were scanned at the beginning of the study and a metric was created. The 3D scanner allowed us to measure the women's bodies more accurately than would have been possible with more limited measurement tools such as tape measures.
* In summary, our hypotheses were as follows:
* - Hypothesis 1: Higher BMI predicts greater body dissatisfaction.
* - Hypothesis 2: The number of retailers carrying one's size predicts body satisfaction and mediates the relationship between BMI and body dissatisfaction.
* - Hypothesis 3: The predictions in Hypothesis 2 will be confirmed regardless of women's age.

## Method

### Design

A prospective design was used in which measurements were taken at two time points. At time 1, women were scanned and their height and weight were measured. At Time 2 (1-3 months later), body satisfaction was recorded as a response to an emailed link to an online questionnaire.

### Participants

#### Participants in this study were all female (N = 85) and aged between 18 and 81 years (mean = 47.81; SE = 1.76). BMI ranged from 18.69 to 56.13 (mean = 26.73, SD = 6.22). In terms of UK National Health Service BMI categories (National Health Service, 2019), 40 (47.1%) were between BMI 18.5 and 24.9 (healthy weight), 27 (31.8%) were between 25 and 29.99 (overweight), and the remainder (21.1%) > 30 (obese). Three women had a BMI greater than 40. Participants self-identified as white (n = 80; 94.1%), mixed white/Asian (1; 1.2%), black (2; 2.3%), Chinese (1; 1.2%), and other ethnic group (1; 1.2%).

### Procedure

## Results

### The number of retailers where participants would find clothing in appropriate sizes ranged from 0 to 39 (mean = 6, SD = 10.50). With an overall rating of 5 (very satisfied), the mean body satisfaction response was 2.64 (SD = 1.24).

### Correlational Analysis

Correlation analysis

Table 1 shows the results of the correlation analysis performed before the mediation analyses. BMI was significantly negatively correlated with both body satisfaction and the number of retailers with clothing sizes that fit. These results support the first steps of mediation (Baron & Kenny, 1986) and show that BMI (predictor variable) is correlated with body satisfaction (outcome variable) and then with British retailers with matching dress sizes (mediator). The proposed covariate, age, did not correlate significantly with the independent, outcome, or mediator variables.

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | | [Table](https://journals.sagepub.com/doi/full/10.1177/0887302X20915528) | **Table 1** Descriptive Statistics and Correlations Between Key Variables. | |

**Table 1** Descriptive Statistics and Correlations Between Key Variables.

Table

Description automatically generated

[**View larger version**](https://journals.sagepub.com/na101/home/literatum/publisher/sage/journals/content/ctra/2021/ctra_39_4/0887302x20915528/20210831/images/large/10.1177_0887302x20915528-table1.jpeg)

### Mediation Analysis

After results showed that BMI was a significant predictor of body satisfaction (β = -.093, SE = .019, p <.05), regression analysis was used to examine whether the availability of appropriately sized clothing in retail stores mediated this effect (see Figure 2).[
                        figure
                    ](https://journals.sagepub.com/doi/full/10.1177/0887302X20915528)

**Figure 2.** Mediation results (without controlling for age).

Results show that BMI was a significant predictor of the number of UK retailers with appropriately sized clothing (β = -.566, SE = .175, p <.05) and that the number of retailers was a significant predictor of body satisfaction (β = .029, SE = .012, p <.05). BMI was also a significant predictor of body satisfaction after controlling for the mediator, number of UK retailers (β = -.076, SE = .020, p <.05). Approximately 27% of the variance in body satisfaction was explained by the predictors (R 2 = .269). To test the indirect effect, a percent bootstrap estimation procedure with 10,000 samples was used. These results showed a significant indirect coefficient (β = -.017, SE = .009, 95% CI [-.039, -.003]). Because the CI did not include zero and the strength of the relationship between the predictor and the outcome was reduced by the inclusion of the mediator, it can be concluded that the presence of more stores where appropriate clothing sizes are available partially mediates the relationship between BMI and body satisfaction.

Figure 3 shows the results of the mediation model when controlling for age. Approximately 27% of the variance in body satisfaction was explained by the predictors in this model (R 2 = .274). The results suggest that age has little effect on the partial mediation of BMI and body satisfaction by the availability of clothing sizes in UK stores.[
                        figure
                    ](https://journals.sagepub.com/doi/full/10.1177/0887302X20915528)

**Figure 3.** Mediation results (controlling for age).

## Conclusions

Having a body size and shape that meets the size specifications of several retailers predicts body satisfaction in women, and BMI predicts body satisfaction independent of age. These correlation data suggest that the lack of availability of well-fitting ready-to-wear clothing in larger sizes may be a key factor in predicting body dissatisfaction among women with higher BMI. This suggests that manufacturers and retailers could improve women's body satisfaction by designing clothing that fits women with larger bodies better and by offering a more extensive range of clothing in larger sizes. However, further research is needed to fully understand the mechanisms behind the relationships observed here.

<https://journals.sagepub.com/doi/full/10.1177/0887302X20915528>

2 FitME:

The goal of this study is to develop a smartphone system that allows users to estimate their body measurements and predict their body size by taking a 2D image of the body with a typical smartphone camera in a specific direction (e.g., from the front). To develop a robust application and train the algorithm with real-world data, an experiment was first conducted to take photos of a set of participants along with their body measurements. Then, the proposed approach extracts the features from the images using computer vision and machine learning techniques to estimate the body measurements (e.g., waist and bust size). Then, the approach uses a Support Vector Machine (SVM) to determine the appropriate size of the shoppers. The use of such a system will help online shopping customers to accurately estimate their body measurements and improve online shopping.

Methodology

Machine vision and learning techniques have been used to estimate people's body measurements and predict their height based on 2D images taken with ordinary smartphones. To estimate the body measurements of a human, the model 1) recognizes the human body in the images, 2) extracts the features of the body from the image, 3) determines the focal points of the human body, and 4) calculates the body measurements by computing the difference between the focal points. To predict the correct clothing size of a particular person, the models use a Support Vector Machine trained on some body measurements and body sizes of customers.Methodology.

Machine vision and learning techniques were used to estimate people's body measurements and predict their body size from 2D images taken with ordinary smartphones. To estimate the body measurements of a human, the model 1) recognizes the human body in the images, 2) extracts the features of the body from the image, 3) determines the focal points of the human body, and 4) calculates the body measurements by computing the difference between the focal points. To predict the correct clothing size of a particular person, the models use a Support Vector Machine trained on some body measurements and body sizes of customers.

3.2 Feature Extraction

Feature extraction is an essential step in image processing that leads to better understanding and classification of images. The main goal of feature extraction in this study is to extract some interesting points (i.e., focal points)

from the detected body to estimate the measurements for each body part (i.e., shoulders, chest, waist, and hips).

The method used in this research is based on two main steps. In the first step, the input image is segmented vertically into parts like. Then, in the second step, two points are determined on the left and right sides of each body part. The selected points should be on the closest segment line to the left and right sides of each body part. For segmentation,it was decided to divide the body image into 40 vertical lines to obtain the line falling on the same or nearby points of interest. To estimate the dimensions of a body part, you need to extract at least two reference points called focal points

(e.g. the left and the right side of the shoulder).

3.4 Estimate the measurements

After you have determined the left and right focal points for all body parts, you must determine the distance between these points by finding the difference between them. The result of the difference represents the width of the area in the unit pixels, but the system works with the unit cm. Therefore,you need to convert the results from pixels to cm. To convert the width of pixels to cm, multiply the difference by 0.0264 according to equation (1). The result you get after converting the difference represents the width of the area within the body image in cm. To estimate the actual size of a body area whenyou know the distance between the focal points in the image, you need to specify a reference point (for example, the ratio between the distance between the focal points in the image and the distance between the real focal points for the fixed distance between the body and the cell phone camera). Thus, we specified the ratio between the width of the focal points in the image (cm) and in reality (cm) (i.e., the ratio between the width of the shoulder in the image and the real measurements of the shoulders, the ratio between the measurements of the chest in the image and in reality, and in the same way the ratio for the waist and hips). The values of the reference points (i.e., the fixed ratios for each body part) are the values of a single randomly selected participant among the men and women. This value is then used to estimate the measurements for all participants. To estimate the measurements of the shoulder, use the measurements of the bust,

𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸 𝑊𝑊𝑊𝑊𝑊𝑊𝑊𝑊𝑊𝑊 (𝑚𝑚𝑚𝑚𝑚𝑚𝑚𝑚) = 26.01 ∗ 𝑤𝑤𝑤𝑤𝑤𝑤𝑤𝑤ℎ(𝑐𝑐𝑐𝑐) (6)

𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸 𝑊𝑊𝑊𝑊𝑊𝑊𝑊𝑊𝑊𝑊 (𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓) = 14.66 ∗ 𝑤𝑤𝑤𝑤𝑤𝑤𝑤𝑤ℎ(𝑐𝑐𝑐𝑐) (7)

𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸 𝐻𝐻𝐻𝐻𝐻𝐻(𝑚𝑚𝑚𝑚𝑚𝑚𝑚𝑚) = 31.15 ∗ 𝑤𝑤𝑤𝑤𝑤𝑤𝑤𝑤ℎ(𝑐𝑐𝑐𝑐) (8)

𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸𝐸 𝐻𝐻𝐻𝐻𝐻𝐻(𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓𝑓) = 39.05 ∗ 𝑤𝑤𝑤𝑤𝑤𝑤𝑤𝑤ℎ(𝑐𝑐𝑐𝑐) (9)

3.5 Predict the size

Developed a machine learning model to predict shopper sizing by training and testing the models on a real-world dataset of participants' body measurements and clothing sizes. In this phase, a supervised machine learning approach was used since the dataset contains features (i.e., body measurements) and labels (i.e., standard clothing sizes). implemented SVM classifiers that can predict the shopper's clothing size. The Support Vector Machine (SVM) is a type of supervised machine learning classification algorithm that was introduced in the 1960s. It is one of the most robust and fastest algorithms among classification methods and produces excellent results. created several SVM models, each of which predicts one of the standard sizes (e.g., XS, S, M...etc.) for each clothing category (i.e., tops, bottoms, and whole pieces) and uses a one-against-all multi-class classification approach. Finally, the output of the models represents the predicted size of the garments.

4. Dataset

The dataset plays an important role in the performance of any machine learning based system. To develop a robust system that helps online shoppers estimate their body measurements from an image taken with a smartphone, a machine learning algorithm must be developed that is trained on a dataset consisting of images of people and their body measurements. The dataset is divided into two parts: the training part and the testing part. The training part is used to train the model and the testing part is used to test and evaluate the model after the training phase. For this study, you will need a dataset that contains images of the human body and basic measurements for each body. The measurements needed for the objective of this research must include shoulder width, chest circumference, waist circumference, hip circumference and some other measurements. Most of the available datasets are either missing basic values such as chest, shoulder and waist or there is no real structured data to help us achieve the main goal of this research. The dataset follows a scientific research procedure. The participants in the dataset are male and female volunteers over the age of 12. We have 34 male participants and 26 female participants. The weight of the male participants ranges from 52 to 103.5 and the length ranges from 151 to 190.

Therefore, we measured the body measurements of the participants manually with a tape measure to determine shoulder, chest, waist, hip, and length in centimeter unit of measurement. In addition, the body weight of each subject was determined with a scale in the unit of measurement kilogram.Sahar Ashmawi et al. / Procedia Computer Science 163 (2019) 209–217 215

Author name / Procedia Computer Science 00 (2019) 000–000 7

5. Results / Discussion

This section presents the results of size prediction using the estimated measurements. Thus, predict the garment size for different garments (i.e., top, bottom, and whole garment) and compare the estimated size to the actual size of the garments worn by the participants.Thus, test the model on a sample of 34 participants, 22 of whom are male and 14 of whom are female. For the first category, shoulder width and chest circumference were used as features to predict the size of the top. The results of the first model to predict top size show that 9 out of 22 males predicted the same true size, i.e., the accuracy of the model is 41%. For the other model, only 3 out of 14 women predicted the same true size, i.e., the accuracy of the model is 21.4% and most of the predicted sizes differ from the true size.

For the second model, waist circumference and hip circumference were used as features to predict the size of the pos. The result shows that for 16 of the men, the size of the buttocks was accurately predicted, with the estimated size identical to the actual size of the lower garments (i.e., the model provided 72.7% accuracy), while only 4 women obtained the true size, which is 28.6% of the model accuracy. For the third model, we used chest circumference, waist circumference, and hip circumference as features for predicting the size of a whole garment (eg,

Dress size), so exclude men from this model.

Table 1. Predict the size of clothes (i.e. upper clothes, lower clothes and full clothes).

ID Predict upper Real upper Predict lower Real lower Predict full Real full

1 M M S S - -

2 M L S M - -

3 M M S M - -

4 XL L S S - -

5 M S S S - -

6 M M S S - -

7 M M S S - -

8 M L S S - -

9 XL S S S - -

10 M S S S - -

11 M L S S - -

12 M M S S - -

13 XL M S S - -

14 M XL S L - -

15 M M S S - -

16 M L S M - -

17 M M S S - -

18 M S S S - -

19 M M S S - -

20 M L S M - -

21 M L S M - -

22 M M S S - -

23 M S M S M M

24 M L M L M L

25 L XS L XS L S

26 L S L S L S

27 M M M M M M

28 L M L M L M

29 M S S S M XS

30 L S S S L S

31 L L S M L L

32 L M L M L M

33 S L S L S L

34 S M S S S M

216 Sahar Ashmawi et al. / Procedia Computer Science 163 (2019) 209–217 8 Author name / Procedia Computer Science 00 (2019) 000–000

Fig. 3. The results of the predictive models.

6.Conclusion

This study proposes an approach that aims to improve and facilitate the online shopping experience by estimating a person's body measurements from 2D images by photographing the body with a smartphone camera. The experiment was conducted with a sample of volunteers who were photographed, manually measured, and asked to provide their actual clothing size to compare the result to the size predicted by the model. For the study, one of the pre-trained computer vision algorithms was used to detect the human body in images. The detectors are designed to detect three parts of the human body: one detector to detect the upper body, another detector to detect the lower body, and the last detector to detect the whole body. After detecting the main body parts, features are extracted by segmenting each image into 40 parts and determining two points as focal points of each body part to estimate shoulder width, chest circumference, waist circumference, and hip circumference. Then, different machine learning models trained on a dataset of measurements are used to predict the dress size as a function of the estimated measurements. Each model was trained to predict the size of a garment (i.e., to predict the size of a top, a bottom, or an entire garment). The results show that most of the predicted sizes have some deviations from the actual measurements of the participants.<https://www.sciencedirect.com/science/article/pii/S1877050919321416/pdf?md5=5dbd63ead7bee387466468dde43aa38d&pid=1-s2.0-S1877050919321416-main.pdf>

Case study on e-commerce platform for clothing – Saenguin

E-COMMERCE

In the e-commerce landscape, we understand it’s important to substantiate the business impact of fit technology. Fit Intelligence provides this information and more. E-commerce teams use:

Online shopping platforms allow customers to store anytime, anywhere without having to go from store to store to find a product or wait in line at the checkout. Despite the advantages over shopping in a store, customers often have concerns when buying products where measurements need to be estimated, such as furniture and clothing. In particular, choosing the wrong clothing size is a common problem faced by many online shoppers. When shopping, customers browse products or goods with the intention of buying them. Shopping is one of the most important necessities of modern life, where people buy things that meet their needs and interests. This process can take various forms, and online shopping is one of them. Especially in recent years, online shopping has taken over the world at a rapid pace. Customers prefer online shopping over shopping ina store because it requires less time and effort. Customers shopping online With the help of online shopping platforms, customers can purchase goods anytime and anywhere without having to go from store to store to find a product or stand in line at the checkout. Despite the advantages over shopping in a store, customers often have concerns about buying products that require estimating measurements, such as furniture and clothing. In particular, choosing the wrong clothing size is a common problem faced by many online shoppers.Especially in recent years, online shopping has taken over the world at a rapid pace. Despite the many advantages of online shopping, some customers may encounter difficulties because they cannot judge the quality of the product, which may result in receiving a wrong, damaged or delayed product. Especially when shopping for clothes, problems of different kinds occur. There are many reasons why they prefer to store in a store rather than online, such as not being able to determine their correct size, try on the clothes, and assess the quality of the material. When determining their correct size, customers either determine their clothing size by manually measuring their body or they choose a size they are used to wearing. Manual measurement is usually done with a tape measure and physical measurement of height, shoulder width, bust, waist and hip area. However, these measurements are not always accurate and appropriate for all types of clothing. Customers may also need to measure additional body parts if they are purchasing more specific clothing (e.g., suits or long-sleeved dresses). For example, if a customer wants to buy a dress or jacket, he or she will need to measure the chest, waist, and hips. For example, to measure the waist area, the customer must measure the waist tightly and completely around the body. When measuring the thigh, the customer must measure the largest part of the thigh. After measuring the body parts, the customer must convert the measurements and choose the right size. Differences in measurements, body parts measured and type of clothing often result in inaccurate manual measurements and cost time. The company sanguin followed the approcch for customer satisfaction with providing reccomendations of the sizes of the clothes , to estimate the right fit of the product with followed the in form of questionnarie approch to estimate the right fit for the cusotmers . As to make it more precise and to predict the right fit followed the approch of machine learning to predcit the approximate measurement with proper height to weight ratio. Thus ,to make the fir more accurate for the customer satisfaction and reduce the number of returns.

What We Do

We harness the power of algorithms and machine learning to give leading brands and retailers all the tools they need to solve sizing, sell smarter, and turn data into actionable insights.

Where We Began

Lack of clarity about fit is the biggest friction on the path to online purchase. Not only does it cripple conversion rates, but it also leads to sky-high return rates that hurt the environment and the bottom line. Having our roots in Germany (where the average return rate for online retailers is a staggering 50-60%), we have been intimately familiar with the extent of the online sizing problem since our inception.

Our initial solution consisted of a web-based questionnaire pop-up modeling service that allowed users to create a personalized fit profile. The technology itself was good, but getting customers to adopt it was a major challenge. In many ways, we faced the same problem as the stores themselves - we wanted customers to solve the sizing problem, rather than do it for them.

This realization quickly led to the development of our flagship product, Saiz, and its underlying data platform, Fit Analytics. More and more of the world's leading brands and retailers are working with us, and the power of our platform continues to grow as we develop the next generation of innovative machine learning solutions.

Shaping innovation to make the e-commerce platform so smart

SAIZ

Saiz is an intuitive sizing advisor that gives shoppers peace of mind and captures important customer information for apparel companies.

Social Sizing

Using advanced machine learning, Saiz matches shoppers with their fit doppelgangers to deliver unbeatably accurate, data-driven sizing recommendations that are easy for end consumers to understand.

Some of the factors that make the saiz the most realiable product:

Height & Weight and Body Dimensions Reports to define size specifications .This allows the creation of products with their customers’ body type in mind thus ensuring higher fit satisfaction.

Age & Gender Reports to collect specific demographic data on their target audience which enables tailored messaging.

Height & Weight Reports to align data to their target audience. These reports can influence model imagery and marketing messaging making both more relatable to their shoppers.

MACHINE LEARNING AND USER EXPERIENCE

How Machine Learning can Drive User Experience

An estimated 91% of customers are more likely to buy something from a company that remembers them and gives them relevant and timely recommendations. For this reason, more and more retailers are using algorithms and data science to match people with items that are a perfect fit for them. This creates a retail environment that is more contextual and gives customers the personalised experience they want: the Saiz tool can make recommendations based on data from the customer's past.

Fit Issue

There is no uniform sizing system in the apparel industry. This has created uncertainty among shoppers when it comes to what items to buy online. An online visitor who is unsure if a particular item will fit him or her may remain a browser and not convert into a customer.

Many retailers underestimate how much fit and size affect their bottom line. However, the long-term consequences of customers purchasing ill-fitting items in a store can lead to negative and costly results for the retailer. These include returns, lack of trust, a poor user experience and, ultimately, diminished loyalty that can force the customer to stop buying items from the store and store with a competing brand.

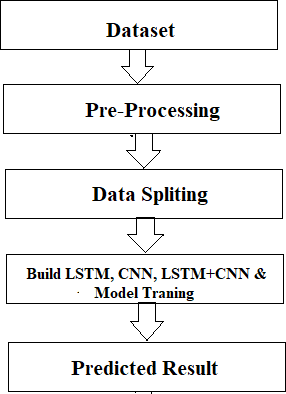
For this reason, apparel e-tailers should choose a solution that increases conversion rates, builds customer loyalty and has a positive impact on sales.

SIZE SOLUTION

SAIZ

The basic tool is SAiz, which gives shoppers recommendations for their ideal size based on a short series of personalized questions such as height, weight and fit preferences. The goal is to use algorithms and a fit-based database to develop an intuitive and sophisticated tool that provides various size recommendations and reduces return rates.

### Proposed Work



*Fig 3.1 Proposed Workflow*

Machine learning is a method of data analysis that automates the creation of analytical models. It is a branch of artificial intelligence based on the idea that systems can learn from data, recognise patterns, and make decisions with minimal human intervention. Importance.

Machine learning is an important way for e-commerce companies to increase sales, meet the needs of their customers, and stand out from the competition. In this article, we explain the benefits of using machine learning technologies.

Machine learning technology is in high demand among apparel retailers, but what is it really?

What is Machine Learning?

In the highly competitive e-commerce space of apparel and footwear, more and more brands are learning how to use machine learning and artificial intelligence (AI) to their advantage.

While AI focuses on a computer's ability to make faster and better decisions that mirror human logic, machine learning is about a computing machine's ability to predict user behavior by analyzing their interactions and behavior.

Artificial intelligence could revolutionize the e-commerce industry because it is a technology that uses experience to improve performance over time.

In other words, the more data that goes through the sophisticated machine learning algorithms, the better the prediction of behavior becomes. This makes it an attractive tool for retailers who want to provide a customer-centric shopping experience.

Key Reasons to Adopt AI in Your E-Tail Business

Offer personalization

Bulk offers are a surefire way to get your customers to look elsewhere.

Machine intelligence is essential for retail companies because of its ability to capture, study and predict user behavior. Behavior, such as which items a customer buys, which products they click on, and the like, are all invaluable pieces of information.

The e-tail apparel industry is a saturated sector. Shoppers are spoiled for choice and they know it. AI technology can help modern shoppers sift through the seemingly endless selection of clothing on the Internet by offering them products that match their preferences.

Increase sales

With the help of machine learning, businesses can offer their customers what they need at the right time. This increases customer loyalty, which can lead to an increase in sales.

When a user has a positive shopping experience with a brand, it increases the likelihood that they will stay on the website longer. It also decreases the likelihood that the user will be tempted to store at other stores.

Tech experts at Outreach believe that "the question of whether artificial intelligence will increase the efficiency and effectiveness of sales teamsis no longer open - it does."

machine learning positively impacts sales by enabling e-tailers to leverage data and make better data-driven decisions.

Machine learning

Reduce returns

Model fit is a measure of how well a machine learning model can generalize to data similar to the data on which it was trained. A model that is well-fitted will produce more accurate results. A model that is too well-fitted fits the data too well.

Every machine learning algorithm has a set of basic parameters that can be changed to improve its accuracy. During the fitting process, run an algorithm on data for which you know the target variable, called "labeled" data, and create a machine learning model. Then compare the results to real, observed values of the target variable to determine its accuracy. Then use this information to adjust the default parameters of the algorithm to reduce the error rate and improve the detection of patterns and relationships between the remaining features and the target variable. Repeat this process until the algorithm finds the optimal parameters that provide valid, practical, and applicable insights for your real-world business problem.

Why is Model Fitting Important?

Model fitting is the be-all and end-all of machine learning. If your model is not properly fitted to the data, the results it produces will not be accurate enough to be useful for practical decision making. A properly fitted model has hyperparameters that capture the complex relationships between known variables and the target variable, allowing it to find relevant insights or make accurate predictions.

Fitting is an automated process that ensures your machine learning models have the individual parameters that are best suited to solve your specific real-world business problem with a high degree of accuracy.

https://www.datarobot.com/wiki/fitting/#:~:text=Model%20fitting%20is%20a%20measure,matches%20the%20data%20too%20closely.

What does fitting a model in machine learning mean?

Let us take an example from regression. Suppose you are given some points (labeled x in the figure below, a relationship between the size of a house and its price).Are to find a model that best represents these points. There are an infinite number of ways to do this. The simplest way is to draw a straight line so that all the points lie on that line. However, as you will discover, this is a poor fit because most of the points do not lie on the line. In this case, it is called an under-fit. Alternatively, you can draw a complex function that fits every point. In this case, it is called an over-fit, which may not generalize to a new point because it has been over- fitted to the given points. The best fit is the middle plot, where you will find a quadratic polynomial that fits the data quite well.

Fitting a model means finding a pattern in data. Let’s say you have a set of points .

x = 1, y = 4

x = 4, y = 10

x = 8, y = 18

Now suppose know that the relationship is linear, so let try to find a model of the form "y = ax + b".

There are infinitely many possible models, and each possible model can be completely defined with two numbers - a and b.

a and b are the parameters of the model.

Parameter learning means that try to find a and b such that the data you have matches y = ax + b as well as possible. If algorithm is good, it should find "a = 2, b = 2" as a solution in this case.

But what if find that the data may not be linear and should also try higher order polynomials?

y = a (P = 0)

y = ax + b (P = 1)

y = ax^2 + bx + c (P = 2)

y = ax^3 + bx^2 + cx + d (P = 3)

…

P is a hyper-parameter. Changing P changes the form of the model.

Statistical Fit

In statistics, a fit refers to how well approximate a target function.

This is good terminology for machine learning, because supervised machine learning algorithms try to approximate the unknown underlying mapping function for the output variables based on the input variables.

In statistics, goodness of fit is often described as referring to measures used to estimate how well the approximation of the function matches the target function.

Some of these methods are useful in machine learning (e.g., computing residual errors), but some of these techniques require that to know the shape of the objective function are approximating, which is not the case in machine learning.

If knew the shape of the objective function, would use it directly to make predictions, rather than trying to learn an approximation from samples of noisy training data.

Overfitting in Machine Learning

Overfitting refers to a model that models the training data too well.

Overfitting occurs when a model learns the details and noise in the training data so well that it negatively affects the model's performance on new data. This means that the noise or random fluctuations in the training data are picked up by the model and learned as concepts. The problem is that these concepts are not applicable to new data and affect the model's ability to generalize.

Overfitting is more likely with nonparametric and nonlinear models because they have more flexibility in learning an objective function. Therefore, many nonparametric machine learning algorithms also include parameters or techniques to limit and constrain how much detail the model learns.

Decision trees, for example, are a nonparametric machine learning algorithm that is very flexible and can lead to overfitting of training data. This problem can be solved by pruning a tree after learning to remove some of the details it has picked up.

Underfitting in Machine Learning

Underfitting refers to a model that can neither model the training data nor generalize to new data.

An underfitted machine learning model is not an appropriate model and is evident by its poor performance on the training data.

Underfitting is often not discussed because it is easy to detect with a good performance metric. The remedy is to move on and try other machine learning algorithms. Nevertheless, it provides a good contrast to the problem of overfitting.

A Good Fit in Machine Learning

Over time, as the algorithm learns, the model's error on the training data decreases and so does the error on the test data set. If you train for too long, the performance on the training data set can continue to decrease because the model makes too many adjustments and learns the irrelevant details and noise in the training data set. At the same time, the error for the test data set starts to increase again as the model's ability to generalize decreases.

How To Limit Overfitting

Both overfitting and underfitting can lead to poor model performance. However, by far the most common problem in applied machine learning is overfitting.

Overfitting is such a problem because evaluating machine learning algorithms on training data is different from the evaluation are actually most interested in, which is how well the algorithm performs on unseen data.

There are two important techniques you can use when evaluating machine learning algorithms to limit overfitting:

Use a resampling technique to estimate model accuracy.

Retain a validation dataset.

The most popular resampling technique is k-fold cross-validation. It allows you to train and test your model k times on different subsets of training data to produce an estimate of machine learning model performance on unseen data.

A validation dataset is simply a subset of training data that withhold from your machine learning algorithms until the end of project. After have selected and tuned your machine learning algorithms based on ytraining dataset, can evaluate the learned models against the validation dataset to get a final objective idea of how the models might perform on unseen data.

What does the word "fit" mean in machine learning?

Usually means how well the "model" that a machine learning algorithm creates (i.e., the algorithm's understanding of the problem, often based on training data) "matches" real-world data/problems.

WhatFirst, should define what a model actually is.

A model is basically a statistical model that describes a relationship between two or more variables. Simply put, machine learning models are mathematical equations of this form:

y=a1x+a2x2+a3

This is a very simple example of what a machine learning model might look like.

Now letus move on to the part about creating these models. These models can be developed by computer scientists and mathematicians by applying the principles of mathematics to data problems. Different methods are used to develop and implement these models and the mathematics behind these models is quite complicated.

Before creating your own model,you should learn about predefined models such as linear regression, support vector machines, and decision tree classifiers that might be appropriate foryour project. A little adaptation of these models could also lead to the desired results!

What problem areyou trying to solve? Discrete or continuous value prediction and estimation offers a wide range of possible tasks. Predicting the stock market or the outside temperature is a continuous value problem, while predicting who will win an election or what fruit will be in a photograph (of a fruit) is a discrete value problem. Predicting ordered discrete values, such as user ratings of movies, is in between.You also need a performance metric that accurately reflects success on that task. First step: choose a task and a performance metric that accurately reflects success.

Data - the type of machine learning model and approach used will depend on what data can be supported.You need a training data set and a test data set that are not highly correlated (to avoid optimistic results). The test data set absolutely needs high-quality ground truth labels and enough samples to adequately assess the performance of each proposed model for the task so that the best possible model can be selected. Ideally, the training data should also have high-quality, unbiased labels and be much larger than the test data set, but this is not always possible. Without tests, you do not know which way the ground is moving, andyou are shooting blind. On the training side, sometimes techniques can be used that are unsupervised or semi-supervised when markers are not available, or even synthesize data. Finally, there is a whole range of techniques for extending days and matching to the domain that can be used to increase the number of training days. All of this really falls under the third step, but it is worth mentioning here. A word of warning - it is generally very difficult to build models without labeled training data, and most of these techniques fail if not enough are available. Second step: define test and training data sets.

Model - which is the best model for the task? This largely depends on whatthe training data can support. Some models work well if you have a lot of data, while others do well with smaller data sets. Each model also comes with a set of design decisions, the most important of which determines the number of parameters to estimate. Pay attention to the relationship between the number of training samples and the number of model parameters! The choice of model goes hand-in-hand with the choice of training algorithm, which includes things like the data augmentation or synthesis mentioned above that can mitigate (but generally solve) a low sample-to-parameter ratio. Third step: train and evaluate the models - "hill climb" to find the best model.

Deep Learning-based solution for recommending appropriate dress sizes.

Basically, the idea is to create a recommendation system for customers based on the product so that bad customer experiences due to products not fitting when purchased can be reduced. This can be due to different sizes across brands and their reporting standards, or not being supplied with the exact product size (which is probably rare), etc.

https://github.com/NeverInAsh/fit-recommendation

Regression in machine learning

Regression in machine learning consists of mathematical methods that allow data scientists to predict a continuous outcome (y) based on the value of one or more predictor variables (x). Linear regression is probably the most popular form of regression analysis because it is easy to use for prediction and forecasting.

5 types of regression in data science

Linear regression.

Polynomial regression.

Logistic regression.

Quantile regression.

Ridge regression.

Lasso regression.

Elastic net regression.

Principal Component Regression (PCR)

Regression algorithms with supervised learning.

What is supervised learning?

Supervised learning is an approach to artificial intelligence (AI) development in which a computer algorithm is trained on input data that has been labeled for a particular output. The model is trained until it can recognize the underlying patterns and relationships between the input data and the output labels, so that it can provide accurate labeling results when faced with data it has never seen before.

Supervised learning is well suited for classification and regression problems, such as determining the category of a news article or predicting sales figures for a specific future date. In supervised learning, the goal is to make sense of the data in the context of a particular problem.

I

How does supervised learning work?

Like all machine learning algorithms, supervised learning is based on training. During the training phase, the system is fed labeled data sets that tell the system what output is associated with each particular input value. The trained model is then fed test data: this is data that has been labeled, but whose labels are unknown to the algorithm. The goal of the test data is to measure how accurately the algorithm performs on unmarked data.<https://www.techtarget.com/searchenterpriseai/definition/supervised-learning#:~:text=Supervised%20learning%20is%20an%20approach,labeled%20for%20a%20particular%20output.&text=In%20supervised%20learning%2C%20the%20aim,context%20of%20a%20specific%20question>.

####################################

Why linear regression?

Simple linear regression analysis is a technique to find the association between two variables. The two variables involved are a dependent variable which response to the change and the independent variable. ... Linear regression is basically fitting a straight line to our dataset so that we can predict future events

Hypothesis

Linear regression is a technique we can use to understand the relationship between one or more predictor variables and a response variable.

If we only have one predictor variable and one response variable, we can use simple linear regression, which uses the following formula to estimate the relationship between the variables:

ŷ = β0 + β1x

where:

ŷ: The estimated response value.

β0: The average value of y when x is zero.

β1: The average change in y associated with a one unit increase in x.

x: The value of the predictor variable.

Simple linear regression uses the following null and alternative hypotheses:

H0: β1 = 0

HA: β1 ≠ 0

The null hypothesis states that the coefficient β1 is equal to zero. In other words, there is no statistically significant relationship between the predictor variable, x, and the response variable, y.

The alternative hypothesis states that β1 is not equal to zero. In other words, there is a statistically significant relationship between x and y.

If we have multiple predictor variables and one response variable, we can use multiple linear regression, which uses the following formula to estimate the relationship between the variables:

ŷ = β0 + β1x1 + β2x2 + … + βkxk

where:

ŷ: The estimated response value.

β0: The average value of y when all predictor variables are equal to zero.

βi: The average change in y associated with a one unit increase in xi.

xi: The value of the predictor variable xi.

Multiple linear regression uses the following null and alternative hypotheses:

H0: β1 = β2 = … = βk = 0

HA: β1 = β2 = … = βk ≠ 0

The null hypothesis states that all coefficients in the model are equal to zero. In other words, none of the predictor variables have a statistically significant relationship with the response variable, y.

The alternative hypothesis states that not every coefficient is simultaneously equal to zero.

The following examples show how to decide to reject or fail to reject the null hypothesis in both simple linear regression and multiple linear regression models.

Example 1: Simple Linear Regression

Suppose a professor would like to use the number of hours studied to predict the exam score that students will receive in his class. He collects data for 20 students and fits a simple linear regression model.

The following screenshot shows the output of the regression model:

The fitted simple linear regression model is:

Exam Score = 67.1617 + 5.2503\*(hours studied)

To determine if there is a statistically significant relationship between hours studied and exam score, we need to analyze the overall F value of the model and the corresponding p-value:

Overall F-Value: 47.9952

P-value: 0.000

Since this p-value is less than .05, we can reject the null hypothesis. In other words, there is a statistically significant relationship between hours studied and exam score received.

Output of simple linear regression in Excel

file:///C:/Users/onkar/Downloads/215761-Article%20Text-531836-1-10-20211009.pdf

<https://www.statology.org/null-hypothesis-for-linear-regression/>

Why is it necessary to clean the data before developing a machine learning model?

As suppose , have not removed unnecessary symbols, punctuation, etc., the system will try to learn from those things as well, which are unnecessary to problem, and will add additional and unnecessary weights/dimensions.

Let say have a word "do not". As humans can understand the semantic meaning of this word. But the model cannot understand it. One solution is called expansion, i.e. convert the word "do not" to "do not" or "can not" to "cannot".

Let assume that taken the data from the Internet. Then there is a good chance that the text in the scraped data contains HTML tags that are also unnecessary for the model that want to create. So should clean those up as well.

There is a saying in the ML community that "Garbage in = > Model = > Garbage out".

This means that if give garbage to the model, it will give garbage as output. In other words: If give the model uncleaned data, it will give bad output.<https://www.quora.com/Why-is-it-necessary-to-clean-the-data-before-developing-a-machine-learning-model>

Selecting the Correct Predictive Modeling Technique

What is Predictive Modeling?

# Predictive modeling involves taking known outcomes and developing a model that can predict values for new events. It uses historical data to predict future events. There are many different types of predictive modeling techniques, including ANOVA, linear regression (ordinary least squares), logistic regression, ridge regression, time series, decision trees, neural networks, and many more. Choosing the right predictive modeling technique at the beginning of project can save a lot of time. Choosing the wrong modeling technique can lead to inaccurate predictions and residuals that have non-constant variance and/or mean.

# Regression Analysis

Regression analysis is used to predict a continuous target variable from one or multiple independent variables. Typically, regression analysis is used with naturally-occurring variables, rather than variables that have been manipulated through experimentation. As stated above, there are many different types of regression, so once we’ve decided regression analysis should be used, **how do we choose which regression technique should be applied?**

<https://towardsdatascience.com/selecting-the-correct-predictive-modeling-technique-ba459c370d59#:~:text=There%20are%20many%20different%20types,neural%20networks%2C%20and%20many%20more>.

What are the advantages of linear regression model?

Point #1. Regression analysis is more versatile and has wide applicability.

Linear regression and Neural networks are both models that you can use to make predictions given some inputs. But beyond making predictions, regression analysis allows you to do many more things which include but is not limited to:

Regression analysis allows you to understand the strength of relationships between variables. Using statistical measurements like R-squared / adjusted R-squared, regression analysis can tell you how much of the total variability in the data is explained by your model.

Regression analysis tells you what predictors in a model are statistically significant and which are not. In simpler terms, if you give a regression model 50 features, you can find out which features are good predictors for the target variable and which aren’t.

Regression analysis can give a confidence interval for each regression coefficient that it estimates. Not only can you estimate a single coefficient for each feature, but you can also get a range of coefficients with a level of confidence (eg. 99% confidence) that the coefficient is in.

and much more…

My point is that there are a bunch of statistical techniques within regression analysis that allow you to answer many more questions than just “Can we predict Y given X(s)?”

Point #2. Regression Analysis is less of a black box and is easier to communicate.

Two important factors that I always consider when choosing a model are how simple it and how interpretable it is.

Why?

A simpler model means it’s easier to communicate how the model itself works and how to interpret the results of a model.

For example, it’s likely that most business users will understand the sum of least squares (i.e. line of best fit) much faster than backpropagation. This is important because businesses are interested in how the underlying logic in a model works — nothing is worse in a business than uncertainty — and a black box is a great synonym for that.

Ultimately, it’s important to understand how the numbers from a model are derived and how they can be interpreted.

Point #3. Learning Regression Analysis will give you a better understanding of statistical inference overall.

Believe it or not, learning regression analysis made me a better coder (Python AND R), a better statistician, and gave me a better understanding of building models overall.

To excite you a little bit more, regression analysis helped me learn the following (not limited to this):

Building simple and multiple regression models

Conducting residual analysis and applying transformations like Box-Cox

Calculating confidence intervals for regression coefficients and residuals

Determining the statistical significance of models and regression coefficients through hypothesis testing

Evaluating models using R squared, MSPE, MAE, MAPE, PM, the list goes on…

Identifying multicollinearity with variance inflation factor (VIF)

Comparing different regression models using the partial F-test

<https://towardsdatascience.com/3-reasons-why-you-should-use-linear-regression-models-instead-of-neural-networks-16820319d644>

1. Archtecture diagram with expaination **Expected Architecture**

**At first , the user will enter the inputs as follows , the model will be able to dectect the inputs and process within the model to ouput the desired measurements.**

INPUTS

OUTPUTS

Machine Learning Model

Backend

FRONTEND

Predicted Measurements displayed

**INPUTS : Height , weight , Body shapes , Age, Gender**

**Expected OUTPUTS (measurement) : Waist , Hip , Chest , Torso Length**

**Frontend : Javascript Framework**

**Backend : Python Flask**

**Machine learning model : Linear regression / Multiple Linear regression**

Methodology

All work flow from data to output therotical

Challenges

Challenges

Although the aforementioned model would work well on shoe dataset, it might not be flexible enough to address the following challenges:

Clothing products like dresses and shirts have relatively more dimensions along which the fit is determined. Furthermore, fit preference might vary across different product categories for each customer, for example, customers might prefer a jacket to be a little loose whereas a wet suit to be more form fitting. Thus, a single latent feature for each child product and customer might not be enough to capture all the variability in the data.

Customers’ fit feedback is unevenly distributed as most transactions are reported as Fit, so it is difficult to learn when the purchase would not be Fit. Standard classifiers are not capable of handling the label imbalance issue and results in biased classification, i.e. in this case, Small and Large classes will have poor estimation rate.

Five Obstacles faced in Linear Regression

This article discusses the problems that may occur while training a Linear model, and some methods to deal with them.

Five problems that lie in the scope of this article are:

Non-Linearity of the response-predictor relationships

Correlation of error terms

A non-constant variance of the error term [Heteroscedasticity]

Collinearity

Outliers and High Leverage Points

Add more if needed

<https://towardsdatascience.com/five-obstacles-faced-in-linear-regression-80fb5c599fbc>

The system presented here composes of five modules:-

* + 1. Input as Dataset
    2. Pre processing
    3. Data splitting
    4. Build & Model train Lstm, CNN and Hybrid approach of LSTM+CNN
    5. Output as Predicted Result

Attribute such as: price of open, high, low, close, adjusted close price taken from huge dataset are fed as input to the models for training to pre-process the data techniques like normalization & one hot encoding in applied on dataset. After this data is divided in two sets namely training & testing which are ratio of 80:20 respectively. Then, this set are used to train a model using 3 different approaches: LSTM, CNN and Hybrid approach of LSTM+CNNS. Finally, all these modules are evaluated using Root mean square error

**Chapter 4. Dataset, Implementation and Result**

### Dataset Detail

The dataset consists of the stock historical data from the National stock exchange (NSE) and captures the daily information of each stock from the National Stock Exchange. It collects different sectors of stock data, including Banking, Pharma, Petroleum, Software and Textiles and it including the opening price, the highest price, the lowest price, the closing price, the adjusted closing price and the volume of stock [18].

|  |  |
| --- | --- |
| **Sector** | **Stock Name** |
| Banking | ICICI Bank |
| Pharma | Sun Pharma |
| Petroleum | GSFC |
| Software | RS Software |
| Textiles | Vardmn Ploy |

*Table 4.1 Dataset Details*

### Tool & Technologies

### PYTHON

The language of select for this project was Python. This was a straightforward call for many reasons.

1. Python [19] as a language has a vast community behind it. Any problems which may be faced is simply resolved with visit to Stack Overflow. Python is the foremost standard language on the positioning that makes it is very straight answer to any question.
2. Python [19] is an abundance of powerful tools ready for scientific computing Packages. The packages like NumPy, Pandas and SciPy area unit freely available and well documented. These Packages will intensely scale back, and variation the code necessary to write a given program. This makes repetition fast.
3. Python is a language as [19] forgiving and permits for the program that appear as if pseudo code. This can be helpful once pseudo code give in tutorial papers should be required and verified. Using python this step is sometimes fairly trivial.

However, Python is [19] not without its errors. The python is dynamically written language and packages are area unit infamous for Duck writing. This may be frustrating once a package technique returns one thing that, for instance, looks like an array instead of being an actual array. Plus the standard Python documentation did not clearly state the return type of a method, this can’t lead without a lot of trials and error testing otherwise happen in a powerfully written language. This is a problem that produces learning to use a replacement Python package or library more difficult than it otherwise may be.

### NUMPY

Numpy is python package which provide scientific and higher level mathematical abstractions wrapped in python. It is [20] the core library for scientific computing, that contains a provide tools for integrating C, strong n-dimensional array object, C++ etc. It is also useful in random number capability, linear algebra etc.

Numpy’s array type augments the Python language with an efficient data structure used for numerical work, e.g., manipulating matrices. Numpy additionally provides basic numerical routines, like tools for locating Eigenvectors

### SCIKIT LEARN

Scikit-learn [21] could be a free machine learning library for Python. It features numerous classification, clustering and regression algorithms like random forests, k-neighbours, support vector machine, and it furthermore supports Python scientific and numerical libraries like SciPy and NumPy.

In Python Scikit-learn is specifically written, with the core algorithms written in Cython to get the performance. Support vector machines are enforced by a Cython wrapper around LIBSVM .i.e., linear support vector machines and logistic regression by a similar wrapper around LIBLINEAR.

### TENSORFLOW

In the TensorFlow [22]has an open source software library for numerical computation using data flow graphs. Inside the graph nodes represent mathematical formulae, the edges of graph represent the multidimensional knowledge arrays (tensors) communicated between them. The versatile architecture permits to deploy the computation to at least one or many GPUs or CPUs in a desktop, mobile device, servers with a single API. TensorFlow was firstly developing by engineers and researchers acting on the Google Brain Team at intervals Google's Machine Intelligence analysis organization for the needs of conducting deep neural networks research and machine learning, but, the system is generally enough to be appropriate in a wide range of alternate domains as well.

Google Brain's second-generation system is TensorFlow. Whereas the reference implementation runs on single devices, TensorFlow can run on multiple GPUs and CPUs. TensorFlow is offered on Windows, macOS, 64-bit Linux and mobile computing platforms together with iOS and Android.

### KERAS

Keras is [23] a high-level neural networks API, it is written in Python and also capable of running on top of the Theano, CNTK, or. TensorFlow. It was developed with attention on enabling quick experimentation. having the ability to travel from plan to result with the smallest amount doable delay is key to doing great research.Keras permits for straightforward and quick prototyping (through user-friendliness, modularity, and extensibility). Supports each recurrent networks and convolutional networks, also as combinations of the 2. Runs seamlessly on GPU and CPU. The library contains numerous implementations of generally used neural network building blocks like optimizers, activation functions, layers, objectives and a number of tools to create operating with text and image data easier. The code is hosted on GitHub, and community support forums embody the GitHub issues page, a Gitter channel and a Slack channel.

### COMPILER OPTION

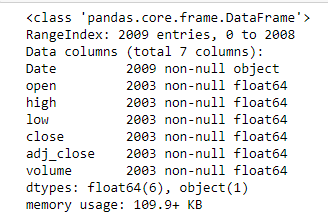
Anaconda is [19] free premium open-source distribution of the R and Python programming languages for scientific computing, predictive analytics, and large-scale process that aim is to modify package managing and deployment. Package versions unit managed by the package management system conda.

### 4.2.7. JUPITER NOTEBOOK

The Jupyter Notebook is an open-source web application that enables to making and sharing documents that contain visualizations, narrative text, live code and equations. Uses include: data , data visualization, data transformation, statistical modelling, machine learning, numerical simulation, data cleaning and much more [24].

* 1. **Results**

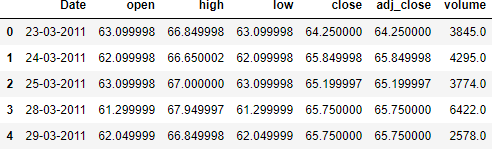
**Step 1:** Dataset Analysis



*Fig 4.1: Stock Dataset Information*

Firstly, I have performed Data analysis for stock price of companies. Fig. represent the date, open, close, high, low, adjusted close and volume of stocks details.

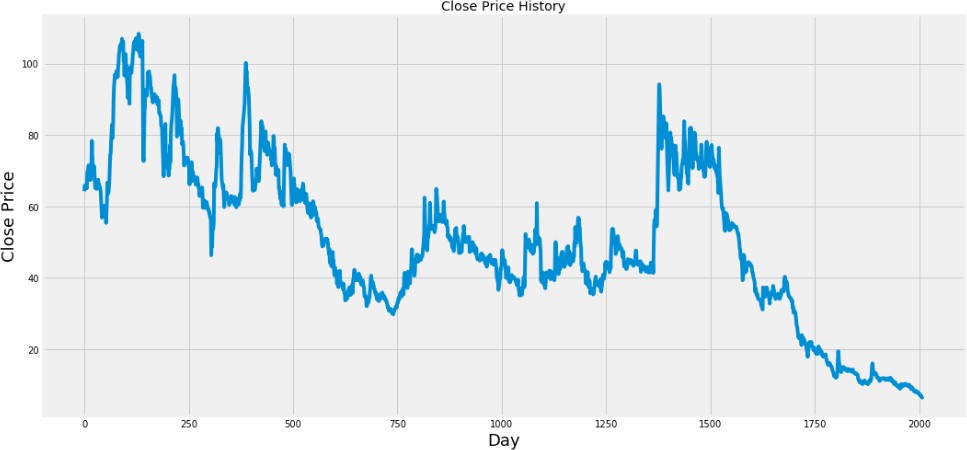
**Step 2:** Read Dataset



*Fig 4.2: Read Dataset*

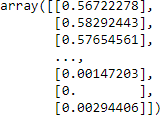
After performing data analysis, I have read the dataset. It shows the dataset information table starting from the tail. There are 4274 data are available in each companies dataset.

**Step 3:** Graph of Close Price history



*Fig 4.3: Graph of Close Price history*

**Step 4:** Preprocessing



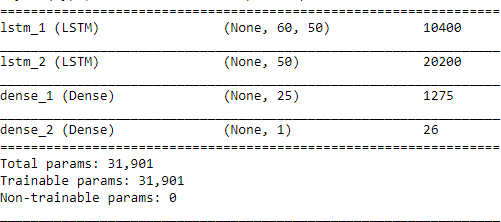
*Fig 4.4: Data Scaling*

After Dataset reading, I have performed preprocessing operation on the dataset. Here I apply Min-Max Scaler to preprocess the dataset. In preprocessing operation removes the noise into the data and convert data into 0 to 1 form.

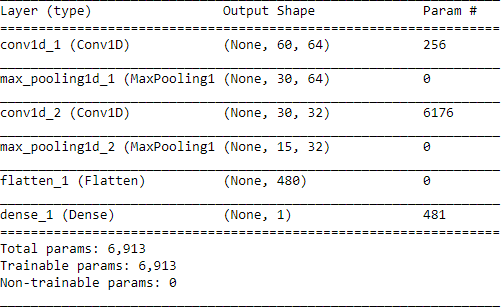
**Step 5:** Train test Split

After performing preprocessing, I have divided the dataset into training and testing set. 80% of the data is used for the training while the remaining 20% of the data is used for testing..

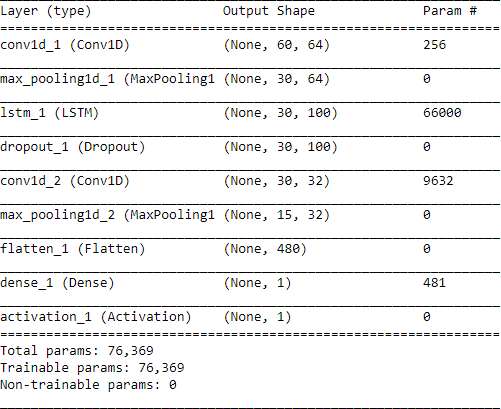
**Step 6:** Model fitting of Long Short Term Memory architecture[25] [26], Convolution Neural Network architecture[27] & Hybride Approach of LSTM+CNN.



*Fig 4.5: LSTM Summary*



*Fig 4.6: CNN Summary*



*Fig 4.7: Hybride Approach of LSTM + CNN Summary*

After generating training dataset, to apply training I have created LSTM, CNN & Hybride Approach of LSTM + CNN network using KERAS. several variations of this architecture using various numbers of layers and various size of Bottleneck layer.

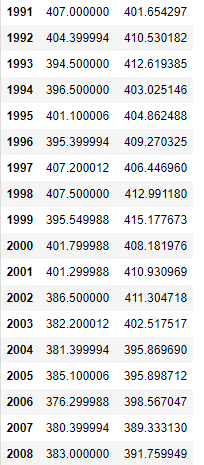
**Step 7:** Apply Training

C:\Users\Janmesh Patel\Pictures\Screenshots\Screenshot (124).png

*Fig 4.8: Training Process*

To apply training, from the samples of Training data, 1543 samples are used for training and 460 samples are used for validation. Data is processed in a batch size of 1 and epoch is 1 for the entire training dataset.

**Step 8:** Predicted Result

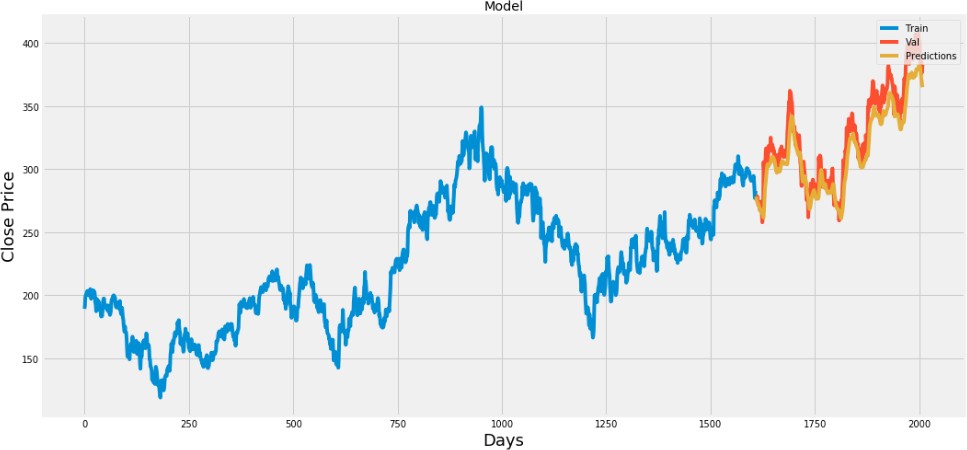


*Fig 4.9: Predicted Close Price*

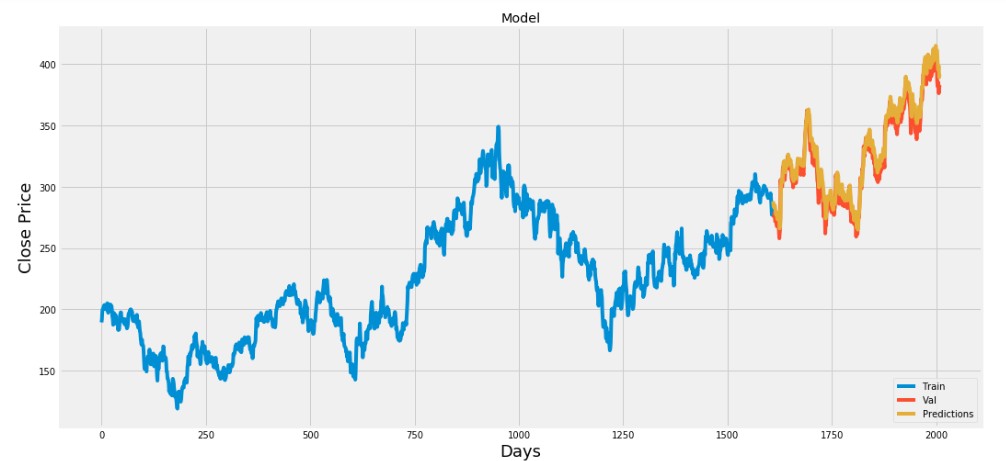
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sector | Stock Name | RMSE | | |
| LSTM | CNN | LSTM+CNN |
| Banking | ICICI Bank | 22.5409 | 8.1499 | 9.1497 |
| Pharma | Sun Pharma | 19.4190 | 16.2015 | 16.0716 |
| Petroleum | GSFC | 5.4396 | 6.6478 | 4.6335 |
| Software | RS Software | 4.7152 | 3.7276 | 3.0633 |
| Textiles | Vardmn Ploy | 1.3909 | 2.5984 | 2.2982 |

*Table 4.2 Accuracy*

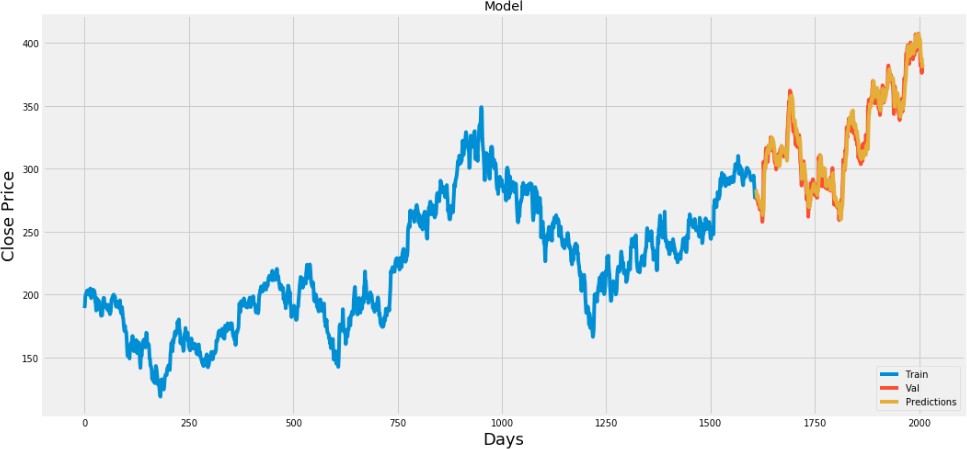
**Step 9:** Predicted Graph



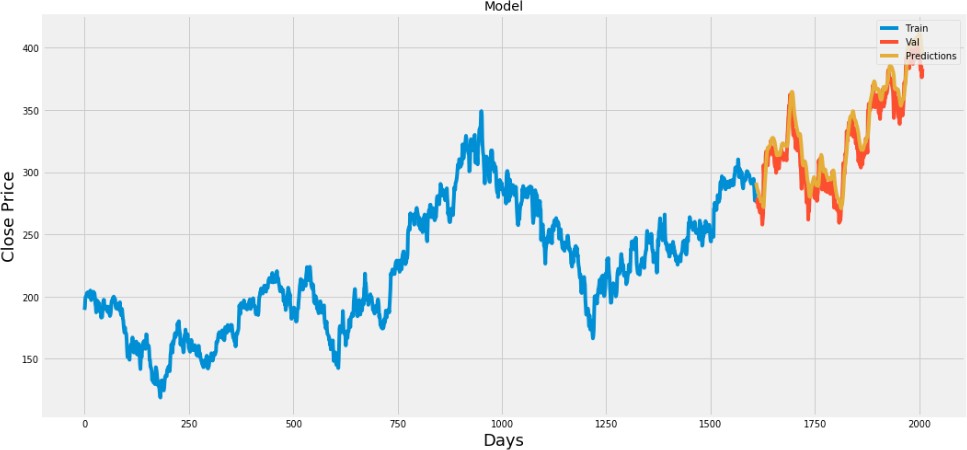
*Fig 4.10: Plot for Real vs Predicted value for ICICI Bank using LSTM*



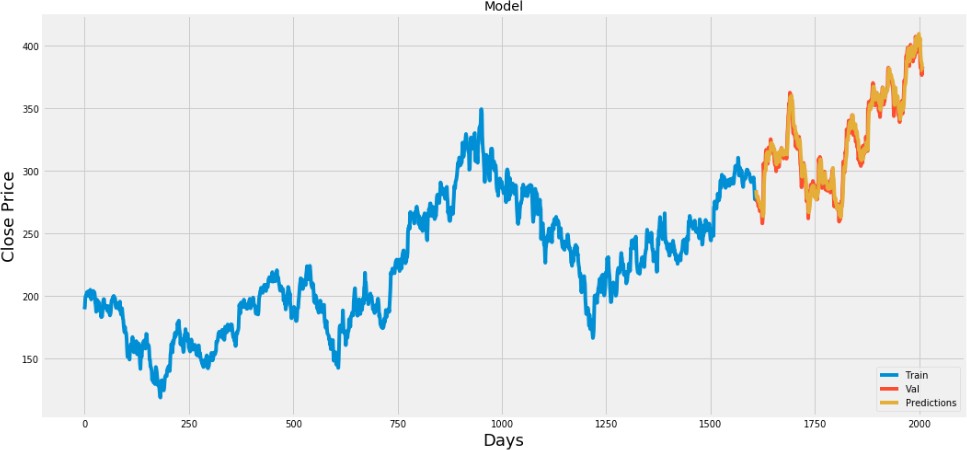
*Fig 4.11: Plot for Real vs Predicted value for ICICI Bank using CNN*



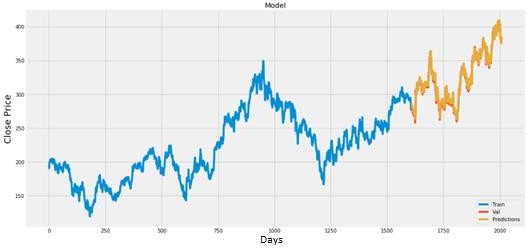
*Fig 4.12: Plot for Real vs Predicted value for ICICI Bank using LSTM + CNN*



*Fig 4.13: Plot for Real vs Predicted value for SUNPHARMA using LSTM*



*Fig 4.14: Plot for Real vs Predicted value for SUNPHARMA using CNN*



*Fig 4.15: Plot for Real vs Predicted value for SUNPHARMA using LSTM + CNN*

*Fig 4.16: Plot for Real vs Predicted value for GSFC using LSTM*

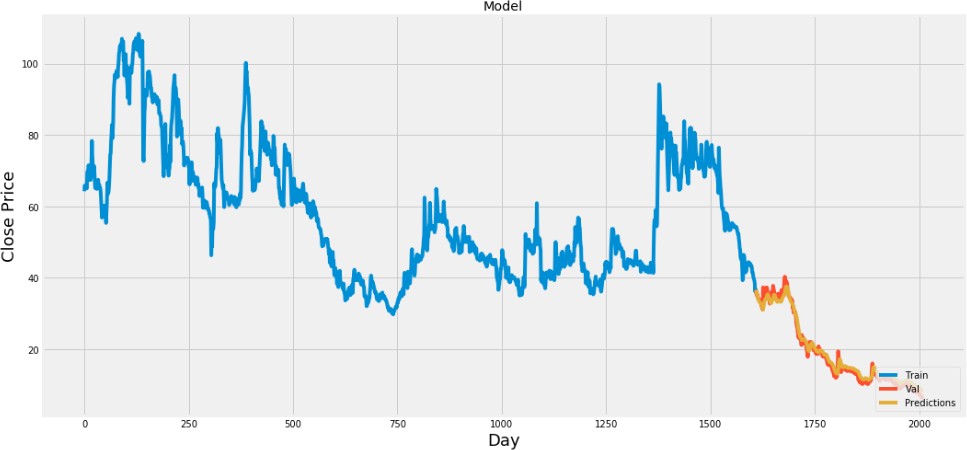
*Fig 4.17: Plot for Real vs Predicted value for GSFC using CNN*

*Fig 4.18: Plot for Real vs Predicted value for GSFC using LSTM + CNN*

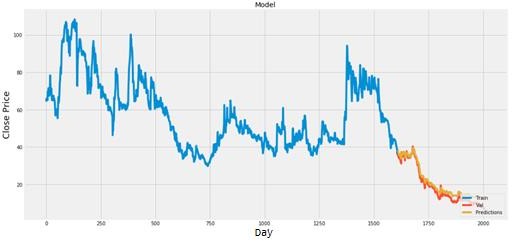
*Fig 4.19: Plot for Real vs Predicted value for RSSOFTWARE using LSTM*

*Fig 4.20: Plot for Real vs Predicted value for RSSOFTWARE using CNN*

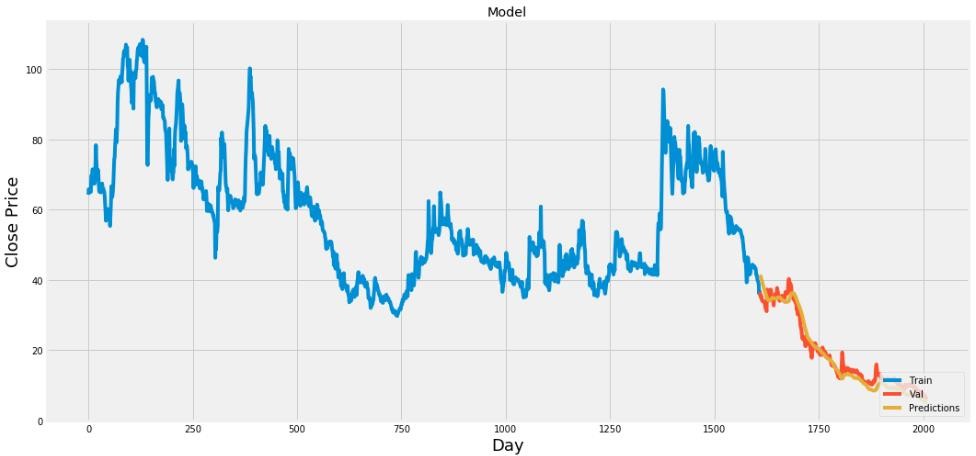
*Fig 4.21: Plot for Real vs Predicted value for RSSOFTWARE using LSTM + CNN*



*Fig 4.22: Plot for Real vs Predicted value for VARDMANPOLY using LSTM*



*Fig 4.23: Plot for Real vs Predicted value for VARDMANPOLY using CNN*



*Fig 4.24: Plot for Real vs Predicted value for VARDMANPOLY using LSTM + CNN=*

**Chapter 5. Conclusion and Future Work**

In report, we will compare a machine learning models like LSTM model, the CNN model and also the hybrid approach of LSTM + CNN model. We have a tendency to train the model using the data of NSE listed companies to predict the stock future value. This is shows the proposed method is capable to distinctive around interrelation with the data. Also, it is evident from the results that, Hybrid approach of LSTM+CNN model is capable to identify the changes in trends. For the projected method the Hybrid approach of LSTM+CNN is known as the best model. It uses the information given at a specific instant for prediction. Even if the other two models LSTM and CNN are utilized in a lot of other time-dependent data analysis, it is not outperforming over the Hybrid approach of LSTM+CNN architecture in this case. This is often because of quick changes occur in stock market. The changes in the stock market is not always be in a regular pattern or not always follow the continuous cycle. Based on the companies and sectors, the existence of the trends and the period of their existence will differ. The analysis of this type of cycles and trends can offer a more profit to the investors. In future work, we add more stock market data and compare more model to improve accuracy of predicted stock price.

In the future, for better accuracy model can be trained with more varied and detailed data. Also, other algorithms along with proposed can be used to create a new hybrid model.

# Bibliography and References

1. “Stock price prediction using LSTM, RNN and CNN-sliding window model - IEEE Conference Publication.” https://ieeexplore.ieee.org/document/8126078 (accessed Dec. 27, 2019).
2. J. Jagwani, M. Gupta, H. Sachdeva, and A. Singhal, “Stock Price Forecasting Using Data from Yahoo Finance and Analysing Seasonal and Nonseasonal Trend,” in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, Jun. 2018, pp. 462–467, doi: 10.1109/ICCONS.2018.8663035.

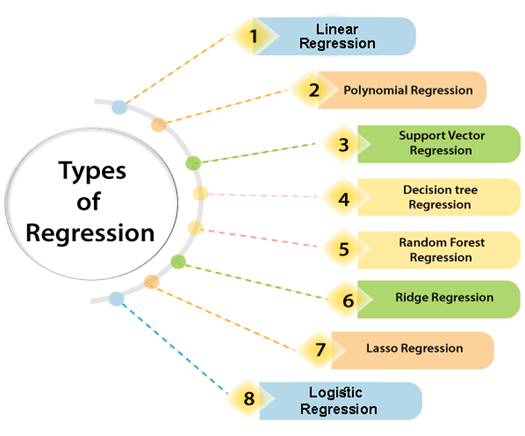
[3] I. Parmar *et al.*, “Stock Dec. 2018, pp. 574–576, doi: 10.1109/ICSCCC.2018.8703332.

1. Y. Lei, K. Zhou, and Y. Liu, “Multi-Category Events Driven Stock Price Trends Prediction,” in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Nanjing, China, Nov. 2018, pp. 497–501, doi: 10.1109/CCIS.2018.8691392.
2. B. Jeevan, E. Naresh, B. P. V. kumar, and P. Kambli, “Share Price Prediction using Machine Learning Technique,” in *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, Bangalore, India, Oct. 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739647.
3. M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali, “Stock market prediction using machine learning techniques,” in *2016 3rd International Conference on computer and Information Sciences (ICCOINS)*, 2016, pp. 322–327.
4. J. Du, Q. Liu, K. Chen, and J. Wang, “Forecasting stock prices in two ways based on LSTM neural network,” in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Mar. 2019, pp. 1083–1086, doi: 10.1109/ITNEC.2019.8729026.
5. S. E. Gao, B. S. Lin, and C.-M. Wang, “Share Price Trend Prediction Using CRNN with LSTM Structure,” in *2018 International Symposium on Computer, Consumer and Control (IS3C)*, Dec. 2018, pp. 10–13, doi: 10.1109/IS3C.2018.00012.
6. T. Gao, Y. Chai, and Y. Liu, “Applying long short term momory neural networks for predicting stock closing price,” in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, Nov. 2017, pp. 575–578, doi: 10.1109/ICSESS.2017.8342981.
7. R. Y. Nivetha and C. Dhaya, “Developing a Prediction Model for Stock Analysis,” in *2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)*, Melmaurvathur, India, Apr. 2017, pp. 1–3, doi: 10.1109/ICTACC.2017.11.
8. Z. Yeze and W. Yiying, “Stock Price Prediction Based on Information Entropy and Artificial Neural Network,” in *2019 5th International Conference on Information Management (ICIM)*, Cambridge, United Kingdom, Mar. 2019, pp. 248–251, doi: 10.1109/INFOMAN.2019.8714662.
9. “Basic LSTM Unit Transfer Function Diagram from [10] | Download Scientific Diagram.” https[://www](http://www.researchgate.net/figure/Basic-LSTM-Unit-Transfer-Function-).[researchgate.net/figure/Basic-LSTM-Unit-Transfer-Function-](http://www.researchgate.net/figure/Basic-LSTM-Unit-Transfer-Function-) Diagram-from-10\_fig8\_308804546 (accessed Jun. 03, 2020).
10. “Fig. 1: An example of CNN architecture.,” *ResearchGate*. https:/[/www](http://www.researchgate.net/figure/An-example-of-CNN-architecture_fig1_320748406).[researchgate.net/figure/An-example-of-CNN-architecture\_fig1\_320748406](http://www.researchgate.net/figure/An-example-of-CNN-architecture_fig1_320748406) (accessed Jun. 03, 2020).
11. Y. Kim, “Convolutional Neural Networks for Sentence Classification,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Oct. 2014, pp. 1746–1751, doi: 10.3115/v1/D14-1181.
12. B. Athiwaratkun and K. Kang, “Feature Representation in Convolutional Neural Networks,” *ArXiv150702313 Cs*, Jul. 2015, Accessed: Apr. 14, 2020. [Online]. Available: [http://arxiv.org/abs/1507.02313.](http://arxiv.org/abs/1507.02313)
13. F. A. Gers, D. Eck, and J. Schmidhuber, “Applying LSTM to Time Series Predictable through Time-Window Approaches,” in *Artificial Neural Networks — ICANN 2001*, Berlin, Heidelberg, 2001, pp. 669–676, doi: 10.1007/3-540-44668-0\_93.
14. J. Han and C. Moraga, “The Influence of the Sigmoid Function Parameters on the Speed of Backpropagation Learning,” in *Proceedings of the International Workshop on Artificial Neural Networks: From Natural to Artificial Neural Computation*, Berlin, Heidelberg, Jun. 1995, pp. 195–201, Accessed: Apr. 13, 2020. [Online].
15. “NSE Listed 1000+ Companies’ Historical Data.” https://kaggle.com/abhishekyana/nse- listed-1384-companies-data (accessed Jul. 07, 2020).
16. A. Tipirisetty, “Stock Price Prediction using Deep Learning,” p. 60.
17. “Python Numpy Tutorial | Learn Numpy Arrays With Examples,” *Edureka*, Jul. 14, 2017. https:/[/www](http://www.edureka.co/blog/python-numpy-tutorial/).[edureka.co/blog/python-numpy-tutorial/](http://www.edureka.co/blog/python-numpy-tutorial/) (accessed Apr. 21, 2020).
18. “Scikit-learn Tutorial: Machine Learning in Python – Dataquest.” https:/[/www](http://www.dataquest.io/blog/sci-kit-learn-tutorial/).[dataquest.io/blog/sci-kit-learn-tutorial/](http://www.dataquest.io/blog/sci-kit-learn-tutorial/) (accessed Apr. 21, 2020).
19. J. Brownlee, “Introduction to the Python Deep Learning Library TensorFlow,” *Machine Learning Mastery*, May 04, 2016. https://machinelearningmastery.com/introduction- python-deep-learning-library-tensorflow/ (accessed Apr. 21, 2020).
20. “Home - Keras Documentation.” https://keras.io/ (accessed Apr. 21, 2020).
21. “Project Jupyter.” https[://www](http://www.jupyter.org/).[jupyter.org](http://www.jupyter.org/) (accessed Apr. 19, 2020).
22. K. Team, “Keras documentation: Search Keras documentation.” https://keras.io/search.html?query=lstm (accessed Jul. 07, 2020).
23. K. Joshi, *kartik-joshi/Stock-predection*. 2020.
24. K. Team, “Keras documentation: The Sequential model.” https://keras.io/guides/sequential\_model/ (accessed Jul. 07, 2020).

Market Prediction Using Machine Learning,” in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*,

# APPENDIX A: ABBRAVIATION

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Abbreviation** | **Meaning** |
| 1 | CNN | Convolutional Neural Network |
| 2 | LSTM | Long Short Term Memory |
| 3 | RNN | Recurrent Neural Network |
| 4 | MLP | Multi-Layer Perceptron |
| 5 | CRNN | Convolutional Recurrent Neural Network |
| 6 | ANN | Artificial Neural Network |
| 7 | KSE | Karachi Stock Exchange |
| 8 | SVM | Support Vector Machine |
| 9 | MSE | Mean Square Error |
| 10 | RMSE | Root Mean Square Error |
| 11 | MA | Moving Average |
| 12 | EMA | Exponential Moving Average |
| 13 | NSE | National Stock Exchange |
| 14 | BSE | Bombay Stock Exchange |
| 15 | P/E ratio | Profit per Earning ratio |
| 16 | ARIMA | Auto Regressive Integrated Moving Average |
| 17 | FEX | Foreign Exchange |
| 18 | SLP | Single Layer Perceptron |
| 19 | RBF | Radial Basis Function |
| 20 | MLR | Multiple Linear Regression |
| 21 | MAE | Mean Absolute Error |
| 22 | GPU | Graphics Processing Units |
| 23 | CPU | Control Processing Units |
| 24 | CSV | Comma Separated Values |
| 25 | GSFC | Gujarat State Fertilizers & Chemicals Limited |
| 26 | NASDAQ | National Association of Securities Dealers Automated  Quotations |

****