Fashion Prediction System

Ratnakaram Sasank
Computer Science of Engineering
Amrita School of Engineering
Bangalore
sasankratnakaram3@gmail.com

Sagala Sai Anvitha
Computer Science of Engineering
Amrita School of Engineering
Bangalore
saianvithasagala@gmail.com

Sagam Prashamsa Reddy
Computer Science of Engineering
Amrita School of Engineering
Bangalore
prashamsareddy02082003@gmail.com

Yerramachu Sunaini
Computer Science of Engineering
Amrita School of Engineering
Bangalore
sunainireddy45@gmail.com

***Abstract* - Patterns in the fashion business move quickly and might last up to eight months. It is a fiercely competitive market where trends are often changing. Knowing the trends and developing accurate projections that consider the always shifting patterns are important when it comes to businesses. Doing so can significantly impact the performance of a company. In order to better understand which tactics should be employed for the greatest result, it is vital to analyze the many strategies provided to obtain a more effective manner of predicting demands and consumer wants.**

Keywords—Fashion, prediction, classifier, regression

I. INTRODUCTION

Fashion is a way of acting or dressing that is popular at a certain moment. It is significant as a way of speaking since it represents the always evolving fashions that various civilizations throughout the world wear. A movement encompassing aesthetic expressions through clothing and accessories could be referred to as fashion. The fashion sector not only deals with garments but also with accessories, belts, and shoes. Numerous reasons contribute to the fact that the fashion market's tendencies are constantly shifting.

Fashion is dependent on current trends and models, which means that trends have a significant impact on the figures and data related to the fashion industry. One-piece swimsuits could be in vogue today but be out of style tomorrow since fashion trends are erratic. When people least anticipate it, fashion trends from decades past may reappear. Because of this, fashion prediction is crucial for businesses to use to ensure they follow the proper trends.

A strategic consideration for producers and merchants, such as the fashion industry, is more accurate demand prediction. For the purpose of introducing the newest trends and supporting the fashion business, fashion predicting has become crucial. Additionally, prediction contributes to the development of strategies, brand recognition, product line expansion, acquisition of new clients, and the effective use of consumer purchasing power. Sales projections in the fashion business are particularly difficult since, in contrast to traditional items, fashion products are characterized by short shelf lives and extremely unpredictable demand due to changing consumer preferences and cyclical fashion trends.

II. LITERATURE SURVEY

The literature discusses several techniques for coping with the challenges that the fashion industry faces when it comes to precise forecasting. These works primarily concentrate on two distinct aspects. In the first strategy, new technologies and statistical techniques are combined to techniques as well as comparisons of how well they operate.

The second research stream is more focused with the proper blending of expert opinion and statistical projections (Davydenko & Fildes, 2013)[1]. For instance, they talk about issues like the benefit of including expert information (Franses & Legerstee, 2013)[2]. Furthermore, they investigate the incorporation of expert knowledge from the past for statistical model adjustment purposes and found that when the model performs poorly, expert judgement produces predictions that are more accurate.

There have been several novel features offered for forecasting automation, such as the addition of prior sales data and past judgements (Franses & Legerstee, 2013)[2]. The first specified research stream will be primarily examined in this study. Lawrence, Goodwin, O'Connor, and nkal are suggested readings for a thorough summary of the critical forecasting research (2006).

Recent research in the first area has concentrated on artificial neural networks (ANNs) for sales forecasting and reports superior performance over conventional ways (Sun, Choi, Au, & Yu, 2008)[3]. Sun et al. therefore present the use of the so-called extreme machine learning (ELM) for predicting the sales of fashion. Due to the ELM's advantages over the traditional gradient-based learning algorithm, including faster learning, Zhu, Qin, Suganthan, and Huang (2005)[4] proposed it. For forecasting in the fashion retail industry, several studies use evolving neural networks (ENN). They perform actual data analysis, and the outcomes are impressive, particularly for noisy data (Au, Choi, & Yu, 2008). The ELM serves as the foundation for Wong and Guo's (2010) hybrid intelligent model, which they propose for mid-term projections.

A model with two automatic systems is presented by Thomassey and Happiette in their works from 2007 and 2010, respectively (mean- and short-term forecasting). They also suggest applying soft computing techniques like fuzzy inference systems and neural networks to address the peculiarities of the garment sector, such as the absence of historical data. This method successfully overcomes the difficulties and yields encouraging outcomes (Thomassey, 2010). But according to the author, garment corporations are unlikely to adopt these strategies (Thomassey, 2014). However, many commercial software products frequently use these methodologies for their forecasts (Jain, 2007), even if the majority of sales gurus merely use these forecasts as a starting point for their own assessments.

In contrast to the models mentioned, Mostard, Teunter, and de Koster (2011) present a novel approach in their work. They concentrate on pre-order demand data and use the preview division, equal division, and top-flop division forecasting techniques. Their research reveals optimistic predictions for each of the three strategies, with the top-flop approach producing the most reliable results.

III. PROBLEM STATEMENT

By gathering, analyzing, and providing actionable insights from data, data science is transforming the fashion industry and has the ability to completely transform it. That's a lot of information, though - like how many people engaged with your social media postings, Instagram trends, what celebrities are wearing, what people are buying, and much more! Therefore, all of this data is mined using cutting-edge models, and the results are transformed into language that even a layperson can understand. With the aid of numerous prediction algorithms from data science, the fashion sector is able to make better business judgements.

IV. PROPOSED MODEL

Our dataset consists of 10 classes through which fashion images can be classified into their respective classes.

Output Labels:

Each training sample and test sample are assigned to one of the following classes -

0 T-shirt/top

1 Trouser

2 Pullover

3 Dress

4 Coat

5 Sandal

6 Shirt

7 Sneaker

8 Bag

9 Ankle boot

A. DATA PREPROCESSING

Data processing is the process of collecting training dataset and turning it into useful information. A readable format is then used to convey the raw data once it has been gathered, sorted, processed, evaluated, and stored. A group of data scientists and data engineers at an organisation usually carry it out in a step-by-step manner. Either automatically or manually, data is processed. The majority of data processing today is done automatically with a computer, which is quicker and produces more precise results. Data can be transformed into many formats as a result. It can be both aural and graphic. It relies on the software being used and the data processing techniques.

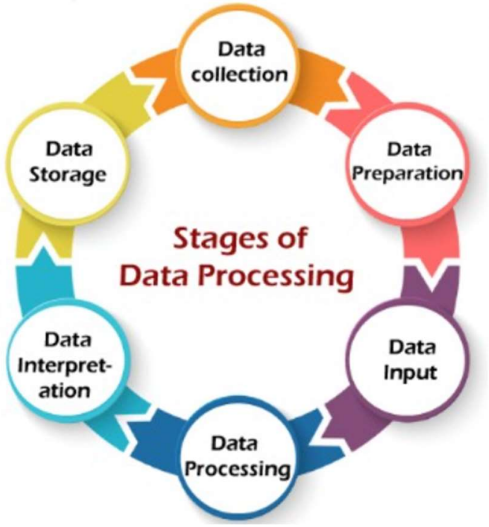

Figure 1: Data preprocessing

a) DATA CLEANING

A vital step in data mining is data cleaning. It plays a significant role in the construction of a model. Although everyone frequently ignores it, data cleaning might be seen as the necessary step. The main challenge in managing quality information is data quality. In information systems, data quality issues can happen everywhere. Data cleaning is the answer to these issues.

Correcting or deleting inaccurate, damaged, improperly structured, duplicate, or incomplete data from a dataset is known as data cleaning. Even though they may appear to be correct, bad data makes outcomes and algorithms unreliable. There are numerous potential for data duplication or labelling errors when merging various data sources.

Data cleaning includes missing data and handling outliers. There are numerous meanings for a missing value. The field may not have been applicable, the event may not have occurred, or the data may not have been available. It's possible that the individual who entered the data didn't know the correct value or didn't care whether a field was left blank. Finding data items with unique characteristics and behaviours from the rest of the cluster or data sets is known as outlier detection. Finding the outliers from the typical items is the process of outlier detection.

b) DATA TRANSFORMATION

To produce patterns that are simpler to grasp, data transformation is a crucial data preprocessing technique that must be applied to the data before data mining.

Data transformation transforms the data into clean, useable data by altering its format, structure, or values. In two steps of the data pipeline for data analytics projects, data can be modified. Data transformation is the middle phase of an ETL (extract, transform, and load) process, which is commonly used by businesses with on-premises data warehouses. The majority of businesses now increase their compute and storage resources with latency measured in seconds or minutes by using cloud-based data warehouses. Organizations can import raw data into the data warehouse without preload transformations thanks to the cloud platform's scalability, then transform it as needed.

c) DATA REDUCTION

Data reduction strategies protect data integrity while minimising the amount of data. Data reduction is the process of taking the original volume of data and reducing it to a much smaller volume. By preserving the integrity of the original data, data reduction techniques are utilised to create a smaller-volume reduced version of the dataset. The effectiveness of the data mining procedure is increased by reducing the amount of data, which yields same analytical outcomes.

Data minimization has no impact on the outcome of data mining. That indicates that the outcome of data mining, both prior to and during data reduction, is identical or nearly identical.

B. DATA VISUALISATION

Data visualisation uses visual components like graphs, charts, and maps to graphically portray quantitative information and data. Data visualisation turns both huge and small data sets into graphics that are simple for people to comprehend and process. Data outliers, patterns, and trends can be easily understood with data visualisation tools. The tools and technology for data visualisation are essential in the realm of big data because they allow for the analysis of enormous amounts of data.

Distribution between two classes is plotted by scatter plot. The link between various variables is displayed using a scatter plot, a sort of data visualisation. Placement of numerous data points between an x- and y-axis is used to display this data. This kind of data visualisation gets its name from the way that each of these data points appears to be "scattered" throughout the graph.

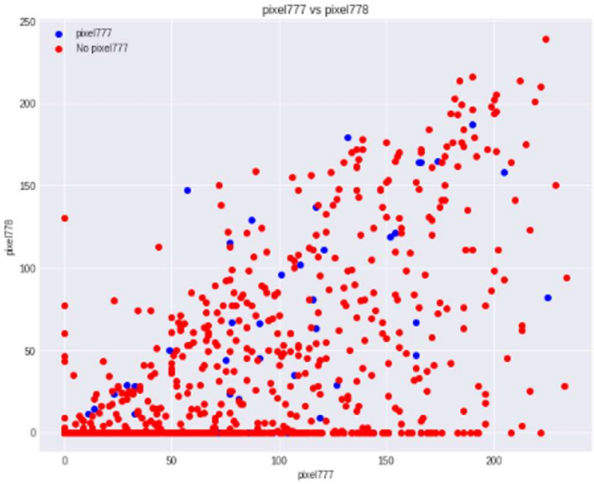

Figure 2: scatter plot

Data distribution of all 10 classes are plotted using strip plot. To summarise a univariate data set, a strip plot is a graphic data analysis approach. The horizontal axis represents the response variable's value, and the vertical axis has all values set to 1. That is, a strip plot is just a display of the answer values that have been sorted along a single axis.

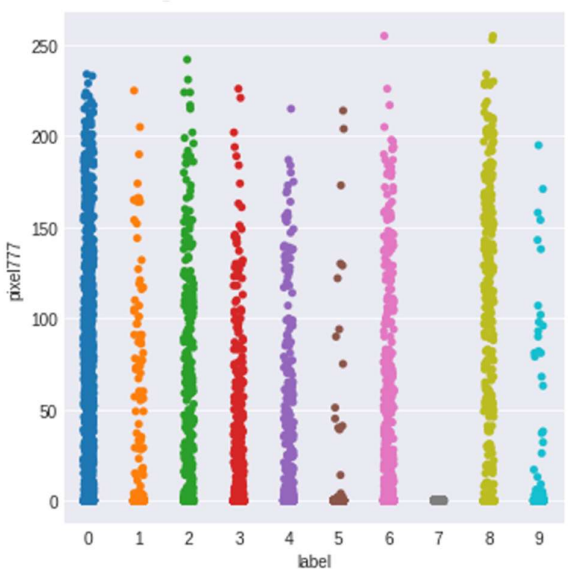

Figure 3: Strip plot

Not only these two plots, based on the data distribution we need we can plot the graphs.

C. SAMPLING

A statistical analysis approach called data sampling is used to choose, alter, and study a representative selection of data points in order to spot patterns and trends in the broader data set being reviewed.

a) Simple Random Sampling

A simple random sample is a population subset that has been chosen at random. With this sampling technique, every person in the population has the exact same probability of getting chosen. Since it only includes one random selection and requires little prior population knowledge, this method is the easiest of all the probability sampling techniques. Any research conducted on this sample should have high internal and external validity because randomization was used, and there should be less chance of biases like sampling bias and selection bias.

b) Systematic Sampling

A probability sampling technique called systematic sampling involves researchers choosing people from the community at predetermined intervals (or k). This method will provide you with a representative sample from which you can make conclusions about your target community if the population order is random or random appearing (for example, alphabetical).

c) Clutser Sampling

In cluster sampling, researchers separate a population into clusters, which are smaller groups. They then choose at random from these clusters to create a sample. Large populations, especially those that are widely geographically distributed, are frequently studied using the probability sampling technique known as cluster sampling. As their clusters, researchers frequently choose pre-existing units like cities or schools.

D. CLASSIFIERS

a) K Nearest Neighbour

Both classification and regression issues are addressed by the K Nearest Neighbors (KNN) technique. It maintains a database of all known use cases and categorises newly discovered use cases (or data points) by grouping them into several classifications. Based on how closely the most recent use cases match the ones that are already available, this classification is completed.

b) Gaussian Naïve Bayes

When discussing classification issues, the term "Naive Bayes" refers to a probabilistic machine learning technique based on the Bayesian probability model. The algorithm's essential premise is that the qualities being considered are unrelated to one another and that changing the value of one does not affect changing the value of the other.

$$P(X|Y)=P(Y|X) * P(X) / P(Y) \quad - (1)$$

c) Support Vector Machine

Both regression and classification tasks are completed using support vector machine methods. These are supervised machine learning techniques, where n is the number of features, and where each piece of data is plotted in an n-dimensional space. It is simpler to plot the features because each feature value has a corresponding coordinate value.

d) Linear Regression

The relationship between an input variable (x) and an output variable (y), also known as independent and dependent variables, is shown via linear regression. To further comprehend the algorithm, consider the scenario where you must place a few plastic boxes of various sizes on various shelves according to their respective weights. It is necessary to perform the work without physically weighing the boxes. The weight must instead be estimated based solely on the height, width, and depth of the boxes. Simply said, visual analysis serves as the foundation for the entire task. As a result, the final configuration of the shelves must be created using a mix of visible factors.

e) Multilayer Perceptron

Multilayer perceptron is made up of dense, completely connected layers that may change any input dimension into the desired dimension. A neural network with numerous layers is referred to as a multi-layer perception. When neurons are combined, some neurons' outputs become the inputs for other neurons, forming a neural network

f) Decision Tree

You may see the road map of possible outcomes for a sequence of decisions using a decision tree. It lets businesses to weigh different outcomes and then make an easy choice based on factors that benefit them, like advantages and probabilities.

g) K means Clustering

Clustering tasks are carried out via the distance-based unsupervised machine learning algorithm K-Means. With the help of this approach, you may group datasets into clusters (K clusters), where the data points from one cluster remain homogeneous and the data points from two distinct clusters remain heterogeneous.

V. RESULT

We have performed various algorithms on our dataset for fashion prediction.

Accuracies for various algorithms are:

1)KNN - 0.88
2)SVM - 78.8
3)MLP - 0.80
4)Decision Tree - 1
5)Naïve Bayes -53.6
6)Linear Regression - 21.56
7)Logistic Regression -24.56
8)K-Means Clustering - 12.4

For all the above algorithms, Support vector machine is the one for which our dataset will give the better accuracy among all the other classifiers

After analysing and comparing accuracies of all the algorithms mentioned, “SVM” is having more accuracy than others. Hence, we consider Support Vector Machine as our model to predict the given image mapping to the classes from 0-9.

VI. CONCLUSION

With a focus on fashion and new product sales forecasting, this survey report provided a review of recent advancements in the field of sales forecasting. For new items and consumer-focused goods, traditional forecasting techniques have difficulty generating accurate sales data. Uncertain demand, seasonality, product unpredictability, and a lack of historical information can be particularly difficult to manage. Hybrid forecasting models function more accurately when used with the most recent methodologies. This study also included a summary of the literature on the predictive value of user-generated content and search terms. Following several parts of fashion forecasting can be useful for future work: In order to provide accurate forecasts, the majority of the examined works propose sophisticated models, and they often validate them using actual sales data.

Investigating models that are both realistic and practical for the day-to-day operations of fashion companies is thus one interesting element. Furthermore, the various affecting aspects that have been mentioned merit further attention. The effects of several factors, including general handling and reduction options, can be looked at. Last but not least, it is important to measure how well user-generated material predicts future events in real life.

ACKNOWLEDGMENT

We would like to thank our course instructors Ms. Priyanka Vivek mam [Data science field], Ms. Shylashree mam [Data Science field] and Ms. Jyotsna C mam [Machine learning field] for supporting us to do this analysis.

REFERENCES

[1] F. Saito, H. Yamashita, H. Sasaki and M. Goto, "Machine Learning for Demand Prediction of Seasonal Second-hand Fashion Items Based on Prior and Fine-tuning Prediction Models," 2021 IEEE 12th International Workshop on Computational Intelligence and Applications (IWCIA), 2021, pp. 1-6, doi: 10.1109/IWCIA52852.2021.9626037.

[2] N. Harale, S. Thomassey and X. Zeng, "Supplier Prediction in Fashion Industry Using Data Mining Technology," 2019 International Conference on Industrial Engineering and Systems Management (IESM), 2019, pp. 1-6, doi: 10.1109/IESM45758.2019.894810