# Classification approach for wristwatch price estimation

regarding the financial benefits of launching a new product

**Katranitsiotis Panagiotis**

2022

# Classification approach for wristwatch price estimation

regarding the financial benefits of launching a new product

by Katranitsiotis Panagiotis

Dissertation submitted to the University of Derby in partial fulfilment of the requirements for the award of Master of Science in Big Data Analytics

2020 - 2022

**Abstract**

It is widely known that modern society encounters numerous daily problems that affect citizens' daily life and routine. No one could underestimate that nowadays the rhythms of life have suffered a remarkable increase. Nevertheless, due to the rapid evolution of technology and computational power, machines and Artificial Intelligence have extreme growth, and by exceeding their applications and capabilities people's daily life have been significantly improved. AI technologies now affect various fields and have become a dominant field in most industries. Besides their constant contribution to the research field, machine and deep learning algorithms can also provide meaningful insights in a business manner. For instance, the most common application of machine learning is the forecasting of a company's annual revenue, and accordingly, through necessary modifications, the most optimal one can be succeeded.

This thesis focuses on an industry-level application of Deep Learning. Particularly, its objective is the creation of an Artificial Intelligent algorithm capable of predicting accurately the price of a wristwatch. Wristwatches industries have suffered substantial evolution throughout the years and due to the growth of technologies, their potentialities became outnumbered. Consequently, their price range has been overextended and has provided the opportunity for all social classes to supply a wristwatch for their daily scheduling. The proposed model will forecast this price by using only visual data, and thus it will provide significant benefits to the wristwatch industry regarding the launching of a new product. More specifically, a designed image could be fed into the model and by considering the predicted price, the company can proceed to a variety of modifications to find the most optimal one. Therefore, it will be able to design the most income-effective wristwatch before the production level and raise its annual revenue. For the efficiency of the model, Amazon's data will be used, since it is the market leader in this category as it provides products in a wide variety of prices and capabilities.

## Declaration

I, Katranitsiotis Panagiotis, confirm that the work presented in this thesis is my own. Information has been derived from other sources and this has been indicated in the thesis.

*Thessaloniki, GR*

Katranitsiotis Panagiotis

## Acknowledgements

I would like to express my utmost appreciation to my supervisor Dr. Gratsia Tantillian for supervising my thesis and my co-supervisor Dr. Berberis for providing guidance and valuable insights throughout the courses of the Big Data Analytics. It has been a great journey, one that I will never forget, and I would like to thank all my professors that supported me during the past two years.

# Contents

## List of Figures



## List of Tables

# 1 Introduction

Over the years people managed to create numerous technologies in order to face their daily problems and improve their daily life and routine. Electricity, heat, communication, and means of transport are among the most significant human inventions. Consequently, since people are dealing with achieving a better way of life, technology has been improved significantly and rapidly. Once people managed to calculate daytime, their daily life was enhanced tremendously both for their agricultural avocations and generally for their routine.

Since the beginning of spoken language, humans have been trying to estimate the time of daily light to handle their daily and weekly schedules. Estimating time started in the ancient years at first with observing the position of the sun and the moon and in later years with simple devices such as hourglass and sundials [1]. No one should underestimate the necessity of them as they led to the monumental breakthrough in technology and the evolution of humanity. However, time device, as we know it in modern society, is strongly related to the wristwatches that most people tend to use.

Nowadays the majority of people using wristwatches to facilitate their daily life. Due to the rapid evolution of technology, wristwatches have been equipped with a variety of capabilities extended their price range. Consequently, most people are able to find a watch that fulfill their needs and budget. Therefore, a statistical analysis as well as Artificial Intelligent technologies could be highly efficient regarding the wristwatches business manner for providing insights and prediction algorithms to manage their annual revenue.

## 1.1 Motivation

It is widely known that nowadays people tend to use smart devices and numerous technologies to plan their daily life and routine. Therefore, numerous data are now stored and are available for scientists and investigators for analyzing them. This leads to the tremendous breakthrough of Artificial Intelligence and Machine Learning Algorithms and Models that are developed in modern society.

Due to the rapid evolution of technology and computational power, a huge amount of data can be processed simultaneously and thus the importance of gathering more and more data has become essential to improve the accuracy of AI models. Most of the devices are connected to the internet and thus all the data gathered and stored in real-time and utilized by companies for beneficial purposes. It has been reckoned that by 2025 more than 150 zettabytes which are equal to 150 trillion gigabytes, will require analysis (IBM, 2020). As a result, most of the problems nowadays are categorized as Big Data ones. Nevertheless, the high volume of data not necessarily defines Big Data but instead numerous factors contribute to this category. These are Volume, Velocity, Variety, Variability, Veracity, Visualization, and Value, known as the 7V's of Big Data [2]. The most significant ones are Volume, Velocity, and Variety. This means that the abundant amount of data, the speed, and the type of data (structured or unstructured) with the combination of the other five characteristics, define the Big Data Problems [3].

Thanks to the variety and complexity of data being available for researchers, machine and deep learning algorithms have increased their accuracy and thus can perform numerous prediction tasks defining AI as we know it in our modern society. Another field of AI that is mainly developed the past few years is computer vision. Deep Learning Models, using visual data such as images and videos, can lead to a extremely accurate results in object, human and animal identification among others.

Consequently, AI models are beneficial to both society and industry level. For instance, every company, regardless of its field, can use machine learning to increase its profit by predicting accurately the financial outcome of the year and even proceeding with some modifications in order to achieve the optimal one. Especially, Computer Vision models, which have numerous identification capabilities and in fact in real time, are highly important for modern society. For instance, this AI technology can be used as security protocols in places, such as airports, to identify a person in real-time without the necessity of human contribution, or on a bank system to ensure that the proper person proceeded to a specific transaction.

For the aforesaid reasons, Machine Learning and generally AI technologies with the combination of the tremendous amount of data which are publicly available face extreme evolution and can contribute to a variety of fields and thus are highly interesting and challenging for research. In addition, they can contribute to a financially beneficial outcome for many companies such as the wristwatch industry.

## 1.2 Thesis Objective

No one could underestimate the utility of wristwatches in a personal manner as they lead to efficient scheduling of daily routine. Nowadays, since most people are familiarized with wearing wristwatches, the technologies they carry, have become outrageous. Apart from providing the current time, many watches have new innovative capabilities such as GPS, heart rate monitor, water resistance, etc. In fact, many of them are Smart Watches and they can also be connected to their user smartphone for handling their calls and calendar. Therefore, wearing a wristwatch is almost essential for effective daily programming. Rolex, which is one of the most popular market leaders in wristwatch production, have sold more than 1,050,000 watches in 2021 [4].

The plethora of wristwatches produced annually, with the combination of the surge of Big Data and the motivation described in the previous chapter, lead to the fact that AI technologies will be purely beneficial and helpful to these industries. The scope of this thesis is to study Machine Learning and Deep Learning Algorithms in a beneficial financially manner for these companies. More specifically, the purpose is to create a model, which uses visual data such as an image of a wristwatch and to be able to predict its probable price. For instance, a company that designs a new wristwatch can use a 3D image design of the watch and will have the ability to forecast its selling price prior to the production process. In addition, in case the predicted price is not satisfactory for the company, the product can be re-designed and fed again to the model to get a new selling price. Consequently, the industry will be capable of achieving the optimal design of the wristwatch product and getting its highest selling price.

To fulfill this goal, extensive research on the types of prediction models will be performed in the following chapters. The purpose of this thesis will be treated as a classification problem by training a Deep Neural Network – a deep learning algorithm - with the aim of matching the new wristwatch to the appropriate class which corresponds to a specific range of price. The extensive research of this procedure and the most suitable classification method will be explained in the next chapters.

Regardless of the method, one of the most vitally important factors for an efficient and accurate model is a sufficient amount of data, as more data lead to the highest accuracy of the model. The data used for this thesis are refer to wristwatches and are going to be retrieved from one of the most famous selling websites, Amazon. The procedure of retrieving the above data as well as the way that they are going to be processed and stored will be explained extensively in the next chapters. More specifically, data that refer to wristwatch characteristics such as color, bandwidth, material, and the price will be gathered along with their corresponding images. In addition, the ethical perspective of the data will be also explained as no personal data were used and all the data regarding wristwatches that were publicly available on the official website of Amazon.

A variety of preprocessed steps and statistical analysis will be conducted on the collected data in order to get helpful insights and relations among the watches' characteristics and their selling price. Furthermore, a clustering methodology and more specifically Hierarchical Clustering will be conducted on the data given these characteristics with the aim to categorize them into classes. Then, a Convolution Neural Network will be trained on the images of the dataset in order to be able to identify the class of a wristwatch given only its image. Consequently, the class categorization of the new product will lead to price prediction as each class will correspond to an average selling price.

## 1.3 Research Questions

In this thesis, the Literature Review will involve the technical perspective for the Machine Learning, Artificial Intelligence along with all the techniques used in this proposed framework as well as the related work regarding price prediction tasks. A statistical analysis of the collected wristwatches will be provided as well as the Deep Learning Model for price estimation will also be presented. The proposed work will be done using

More specifically, the research questions of this thesis are summarized as follows:

1. Which are the characteristics that mostly affects the price of the product and how can they be used to create the appropriate clusters?

2. What is the optimal number of classes for wristwatches to be categorized regarding their characteristics?

3. Which model is most appropriate to identify the class of a new wristwatch?

4. What is the optimal number for the training set and how to evaluate the created model regarding its accuracy?

5. Which parameterization provides the highest accuracy on price prediction

Therefore, in the following pages, the extensive framework for answering the above research concerns, will be presented extensively followed by the conclusion related to the model's result and accuracy. The thesis work will be provided in `Python`

# 2 Literature Review

## 2.1 Big Data

Nowadays more and more people tend to be familiarized with smart technologies so as to improve their daily life. Due to the rapid evolution of internet speed, and computer hardware, no one could deny that we live in the age of technology. As a result, there is a tremendous amount of data produced daily. In previous years, despite the lack of various technology, the way to store such amount of data seemed impossible. Even in cases when data could be stored, it was extremely difficult for computers to handle and process them. Considering the hardware being available for a computer, during the previous decade, RAMs, CPUs and GPUs were insufficient technologically to store and particularly to handle such amount of data. As a result, the breakthrough of the technology not only led to massive production of smart devices but also to the huge amount of data.

Big Data is a term regarding the massive amount of data along with their complex structure and they form the modern field of Data Analysis and Science as well as the field of Artificial Intelligence [5]. Due to the expansion of storing and generating data, the field of Big Data Analytics consist of many challenges and has become one of the most interesting researching areas for scientists and developers [6]. Many companies and industries from different fields have turned their attention by investing in this domain for profitable purposes.

Except from smartphones and there are plenty of devices that store data capable of a beneficial analysis. In modern homes, most of the devices such as computer, TV, air-condition, coffee makers are produced in a way that store data, so the customer can handle them in an easier and more convenient way. Furthermore, almost all the applications installed on mobile phones, computers, laptops, tablets, etc. even websites store information to offer a better personalized experience.

Big Data can get various forms. Particularly, apart from uncomplicated numbers, image, video, and sound data, can now be able to be stored and processed. Images consist of pixels, are represented in an array with the shape of the dimension of the image. Each element of the array contains the value of the color of a specific channel. Colored images contain three channels, red, green, and blue (RGB). Consequently, colored images are defined in three arrays one per color channel. Taking this representation into account it is understood that a single image even in low quality (small size) contains numerous data and processing them occupies a high capacity of memory space. As far as Videos are concerned, they can be split into frames and thus can be handled as images. Sounds are waves and hence they can correspondingly be represented as numbers. Thanks to the evolution of Big Data these fields now are able to be processed by computers and analyzed by researchers.

Most people are being confused with the term of Big Data as they relate this field with a tremendous amount of data. Nowadays the domain of storing data has received exponential growth that except from Gigabytes there are new with higher capacity measurements of storing such as Zettabytes (ZB) or Yottabytes (YB). In fact, the volume of the data although it is a necessary characteristic for defining this domain, it is not the unique one. Besides high Volume, Big Data is also characterized by high Velocity and Variety, the 3V's as commonly known. Velocity refers to the speed of the data. More specifically, it measures how fast the data come. For instance, some data are given in real-time whereas some of them are generated in batches. Big data fundamentally handle this explosion of the data by accepting their incoming flow. Variety corresponds to the structure of the data. Both structured and unstructured data are generated continuously by humans and computers are basic aspects of Big Data. Structured data are numbers, images, and videos whereas unstructured data are text or information gathered from e-mails, websites, internet posts, etc.

In modern society, since Big Data field has received significant growth, it is now defined by 7 characteristics (7V's). Apart from the three basic characteristics, Variability, Veracity, Visualization, and Value are currently define Big Data. Variability although it is similar to Variety, refers to the homogeneity of the data as the constant change of information has a huge impact on data. Veracity is related to the accuracy and the truthfulness of the data along with the provenance of the source. In particular, it is directed to the trustworthiness of the sources that generate data in a constant manner [7]. The ability to visualize Big Data using charts and complicated graphs is also essential as well as acquiring the correct and proper value of information as processing them requires a significant amount of time period and intricate computations.

Big Data Analytics has succeeded to contribute to a variety of fields both scientifically and industrially. They have contributed to the evolution of Data Analysis and Data Science as we know them nowadays. Therefore, they form the field of Artificial Intelligence since they are the "flagship" of Machine and Deep Learning which are the most dominant and highly evolving era of research and are going to be investigated extensively in the following chapter.

## 2.2 Artificial Intelligence

Artificial Intelligence (AI) is the field that focuses on simulating human intelligence in machines and computers [8]. In fact, it aims at learning and not at guidance. Precisely, its purpose is to make algorithms and models for computers to decide to perform a specific task by learning it, rather than being programmed to do so. Therefore, the model will be able to adapt to new environments and circumstances and thus can generalize its performance as people tend to do. Most people relate AI to robots as they are more familiar with these technologies through movies, series, and video games. However, AI refers to every machine capable of performing multiple and complex computational tasks as each command and information that the model receives, is translated into numbers and mathematical operations.

Since computers have enhanced their capabilities and potentialities, the time of performing complicated mathematical equations and operations has been reduced significantly. Making an AI model possible of "thinking" and "deciding" efficiently and accurately depends on its capability of performing fast math operations. Thus, Artificial Intelligence has become one of the most dominant areas in the recent area of research. The breakthrough of AI has led to the invention of new technologies as long as with the development of the existing ones. Industries, companies, devices, and research centers use AI to a great extent or at least on partially level.

Thanks to Artificial Intelligence various fields have been evolved. Autonomous driving now is not only a scenario written for video games and movie series, but it is in point of time and in fact a productivity level. Furthermore, person and object detection are also based on Artificial Intelligent algorithms. Models have been created to be able to detect highly accurate a person and generally an object only with visual data and even in real-time. The most common model with this aim is `YOLO` (You Only Look Once) which real-time object prediction. The combination of high speed and accuracy made this model extremely useful and a State-of-the-art algorithm because it can contribute to security purposes (airports, banks, etc.), vehicle tracking, object detection, and others. More specifically, its efficiency lies in its structure. YOLO's approach is to apply the model on different scales and locations to find the corresponding detections [9]. Other models with high accuracy and similar aims are Fast R-CNN and Mask R-CNN. Both models are extremely complex and although they led to high accuracy, they are not preferable for real-time productions due to their computational speed [10].

Another illustration of the necessity of AI is self-driving cars. Nowadays numerous cars have the ability to understand their surrounding environment, through a lot of cameras and sensors deployed in the vehicle, and thus they able to perform the appropriate tasks. For instance, the existence of an upcoming vehicle, the road signs, traffic lights, etc., and hence it can succeed the autonomous driving. Without AI this will not be achievable as there is no possibility to program all the potential combinations of vehicles, signs, traffic, and weather conditions in order to succeed the autonomous driving. On the other hand, researchers needed to create an AI system in which the car will be able to think and adapt to road conditions and can proceed with the appropriate decisions, such as turning right, or pressing the break, just like humans would do. The most common AI system for the above actions is Reinforcement Learning which is a Deep Neural Network with the intent to create a model which succeeds in performing tasks by a learning method, the teacher-student algorithm as scientists name it [11]. This algorithm is based on rewarding the AI model when conducting a task correctly (such as parking a car in the aforesaid paradigm) and punishing the undesired behaviors [12]. All the above are translated mathematically, so as to be processed by a computer and performed on simulating environments until they succeed the optimal accuracy and subsequently, the model is set on the corresponding vehicles. The definition of

Neural Networks (NNs) and Deep Neural Networks (DNNs) as well as their necessity, their function, and the way they are trained will be explained extensively below since the thesis methodology will follow the approach of a DNN to achieve its objective.

Artificial Intelligence and its subsets of Machine and Deep Learning affect our society to a great extent. Its capability for valid predictions is strongly related to the development of companies. Nowadays, industries can predict their annual revenue using AI models and accordingly can proceed with several modifications in order to find the most optimal ones. Both small and large companies have currently Analytics and AI departments so to be able to conduct statistical results concerning the factors that impact the market and their products and hence find business solutions regarding their revenue.

## 2.3 Machine Learning

Machine Learning (ML) is a type of Artificial Intelligence that allows machines to conduct accurate predictions and outcomes by a specific input, without being programmed to do so. More specifically, the ML models learn through data, identify patterns and following complex processing, learn to forecast with accurate results [13]. Regardless of the type of data, machine learning has been evolved into processing different types simultaneously and thanks to Big Data, more data are available and thus the predicting models have enhanced their efficiency tremendously.

The importance of Machine Learning is remarkable and led to the establishment of Artificial Intelligence. ML models are valuable in numerous fields such as fraud detection, spam filtering, financial predictions, and many powerful applications in medicine. Leading companies such as Facebook, Google, and Microsoft have their central operations surrounding on Machine Learning and AI, to succeed their dominance among the competitive industries [14]. The high volume of data, computational and internet power has enhanced the development of the models and consequently, their efficiency underwent a great growth [15].

### 2.3.1 Data Split (Training and Test Sets)

Machine Learning models along with the data analytics have evolved tremendously in the past years. The high volume of data is one of the basic factors for their development. Since machine learning models are being trained to perform accurate predictions, a remarkable amount of data is essential. The more data being available for processing, the more accurate forecasts the model can result in. However, the huge volume of data does not necessarily lead to the optimal accuracy and efficiency.

AI models use a remarkable amount of data for the training procedure. Nevertheless, they need another set of data for testing their predictions so the researchers can calculate their accuracy of them. More specifically, the given dataset needs to be split into two sets, the training and the test one. It is common for the training set to consist of 70% of the data and the remaining 30% to be the test set. The model uses the training set to learn the relations among the data and to identify patterns to generalize in new unlabeled data. As the model will be created, the test set is going to be used for evaluation. Since the test set contains the true values of the parameters (ground truth), the data are provided into the created model, and by comparing the predictions and the actual values, the accuracy is calculated.

The most common statistical criterion for a model's error is Mean Squared Error (MSE). The MSE measures the average squares of errors and is used to evaluate the estimators given by the output of the model and the actual values. Given the test set as input to an AI model, the actual values are known and by accounting the predicting values, the MSE is calculated. More specifically, it measures the closeness of the estimator to the true value and reflects both the bias of the estimator (the deviation from the actual value) and its variance (precision) [16]. Consequently, the goal of each model is to reduce this error to its minimum level.

As far as the correlation between MSE and the volume of training data, it is common that more data result in higher accuracy. However, there are several cases in which the high amount of data does not necessarily reduce the error of the model. This emerges from overfitting, in which the model learns the training data precisely and thus cannot generalize efficiently its predictions to unknown and unlabeled data. To avoid this condition, the usage of another data set is essential, the validation set. This set does not require to include a lot of data while it is common to be consisting of 10 of the initial dataset. Therefore, the data set should be split into the training (70% of the data), the validation (10% of the data), and the test set (20% of the data). Unlike test set, the validation one will be used during the training procedure so the MSE will be calculated constantly and not just at the end of it. As a result, by observing the value of the error graphically, the optimal model will be appointed to avoid overfitting and eventually provide the most accurate predictions.

The Machine Learning Algorithms are categorized into two basic categories regarding their type of prediction, the classification, and the regression one. Their difference lies on the type of the model's output whether this would be a number (regression) such as financial revenue forecasting or a class (classification) such as gender prediction. Regarding this thesis, its objective is categorized as a classification problem, in which a price class will be predicted given a new unknown wristwatch.

In most forecasting and price prediction problems the actual value of price class is known so the model will be able to be trained to categorize new observations. The innovative about this thesis goal lies on the fact the classes of the wristwatches (ground truth) are unknown and they will be conducted through machine learning.

### 2.3.2 Supervised Learning

The other two basic subsets of Machine Learning problems apart from the type of the predicted output, refer to their learning procedure and they are categorized as Supervised and Unsupervised Learning types.

Regarding supervised learning the goal of the prediction is predefined, and it uses label and annotated datasets for the training procedure. The purpose of machine supervised learning is to classify the data and forecast the desired outcome accurately. In this type of problem, the data have been already labeled and categorized into classes and hence the model has the knowledge, during training, to recognize these specific classes. Consequently, it generalizes its forecasts into unlabeled data and predicts their corresponding classes accurately and efficiently. The number of classes that can predict, is the ones that are already predefined by the labels of the data.

An equivalent strategy is followed on the regression problems as well. Supervised learning, in that case, attempts to understand the relations among the variables (between the dependent and independent ones) [17]. The reason why this problem is categorized as a supervised one, lies in the fact that the relations of the data are already provided and the model, taking advantage of this knowledge, focuses on generalizing its predictions on unknown data. The most common techniques for supervised machine learning are Linear Regression (regression problem) and Logistic Regression (classification problem).

Supervised learning has a plethora of applications in machine learning and their activity in the field of AI is remarkable. They are based on the idea of a supervisor who advises and instructs the model during the learning procedure on what labels should identify [18]. Their dominance is noticed in many multimedia problems as well as in medicine, financial prediction, etc., and in most of the classification problems.

### 2.3.3 Unsupervised Learning

Unsupervised learning utilizes machine learning to identify the relations among the data as well as to analyze and cluster the unlabeled data. For instance, given a dataset in which the observations are not categorized into classes, the unsupervised techniques will exploit and arrange the data into clusters [19]. Cluster is similar to class with the difference that the first one is provided via machine learning (unknown classes) and the second one is given on the initial state of the dataset (labeled data).

Particularly, in unsupervised learning, the model receives an input sequence of data without the knowledge of the target outputs and therefore it tries to identify them itself by learning some patterns for the data environment [20]. These algorithms focus on classifying the data which is recognizing the similar characteristics among them. Translating the above procedure mathematically, the model will calculate the distances among all the observations from the data and will consider the data with a small distance as similar and thus will belong to the same cluster.

The most common techniques for clustering are k-means clustering and hierarchical clustering. Both of them are based on considering the distances among the variables but with a different calculation approach [21]. The most noteworthy distinction between them lies in the number of clusters. On k-means, the k parameter represents the number of classes the model needs to define which has to be given prior to the training procedure. On the other hand, in hierarchical clustering, the number of clusters is not predefined, whereas numerous of them are calculated and provided by the hierarchical tree and then the user considering the most optimal, determines the corresponding clusters that better represent the problem case.

For the objective of this thesis, the number of classes representing the specific dataset of wristwatches is not defined. Consequently, an unsupervised technique will be used for clustering the data and more specifically, the hierarchical clustering approach, so as to find the most optimal amount of them by considering the distances among the variables.

### 2.3.4 Hierarchical Clustering

Hierarchical Clustering is an unsupervised learning technique to define the clusters of the given data. It is based on calculating the distances among the variables and due to their comparison, it is able to categorize the similar data into the same cluster and the dissimilar ones to different ones. The most common metric for this consideration is the Euclidean distance which defines the length of the line between two points in the Euclidean space. However, especially on classification problems there are other metrices that leads to more efficient results into distinguishing the corresponding number of clusters.

It is based in two methods, the agglomerative and the decisive clustering algorithms. The agglomerative approach is going to be implemented in this thesis in which the data points start as individual groupings and then they are merged with their similar, closed distanced variables and thus the clusters are defined. More specifically, it builds a merged tree that contains variable elements to the root and the full data set [22]. Its graphical representation (dendrogram, which is called hierarchical tree) is significantly useful for the optimal pruning of the tree as it also indicates the distances among the data and lead to the definition for the number of classes.

Hierarchical clustering uses linkage methods for its implementation, such as average, ward, single and complete. The most common function, the one that this thesis also follows, is Ward's one. More specifically, this technique in contrast to other existing methods does not measure the distances directly whereas it analyzes the variance of each cluster. Particularly, at every stage, it considers every possible union of the cluster and in the merging procedure, it accounts the minimization of the sum square of loss [23]. The distance between two clusters is defined as the sum of squares which indicates that during merging this number constantly being increased. Ward's method is capable of trying to keep this growth as small as possible and accordingly it is the most suitable approach for defining the efficient number of clusters [24].

### 2.3.5 Principal Component Analysis (PCA)

Prior to each statistical analysis and AI models, the preprocessing of the data is crucial and significantly affect the results and the accuracy of them. In order to provide efficient predictions, the appropriate format and preprocessing of the data is essential. Especially, in cases of unsupervised learning, preprocess techniques can reduce the complexity of the clustering procedure remarkably and in fact lead to a better outcome.

As it was described extensively on the previous chapter, all the unsupervised techniques need to measure the distances among the variables for categorizing them into clusters. Especially, in hierarchical clustering where the number of clusters is not predefined by the user, the complexity is erased exceptionally. For instance, given three variables as input, the model translates this problem as a three dimension one in which it needs to measure all the possible distances among the points, which are the elements of each variable. Considering that in most cases the variables are more than three such as five, six, or a lot more, the dimension of the problem has been increased. This leads to a more complicated approach which not only requires high mathematical operations but also consumes a significant time and space from a computer to handle them.

Reducing the dimensionality of the problem affect incredibly the complexity of the mathematical calculation. However, by reducing the number of variables, the machine learning method deprives the opportunity of generalization as it is not capable of understanding all the relations and patterns among the variables. As a result, the optimal approach is to reduce the dimension of the problem but keeping all the information given by the variables. These methods are the dimensionality reduction ones and the most dominant is the Principal Component Analysis.

Principal Component Analysis is a dimensionality reduction technique, that reduces the complexity of the clustering procedure and contributes to accurate outcomes. More specifically, its aim is to extract the information from the table which contains the values of all the variables and to transfer them into a set of new orthogonal variables which are called principal components [25]. The number of the principal components can be predefined but also can be selected by the algorithm considering a corresponding percent of the information that the user wants to preserve. From a mathematical perspective, PCA uses multiple techniques such as covariance, correlation, and normalization to achieve dimensionality reduction as well as to retain the variability of the information [26]. Last but not least, since it is a distance-based technique, the prior normalization of the data is essential for reducing its complicated operation.

Consequently, after the PCA, the artificial intelligent model will use the new components as the new variables and thus perform the clustering procedure with fewer elements but with the same valuable information. In the case of this thesis, in which the clustering procedure is essential for categorizing the wristwatches, Principal Component Analysis will first be performed to attain more promising results.

### 2.3.6 Neural Networks

There are several types of Machine Learning Algorithms capable of providing extraordinary outcomes such as Support Vector Machines, Linear and Logistic Regression, Decision and Regression Trees among others. A variety of them offer better results on regression or classification problems with some of them to provide useful forecasts on every category. The most dominant machine learning algorithm nowadays are the Artificial Neural Networks (ANNs).

Neural Networks are based on Artificial Intelligence technologies in order to create accurate forecasting models. They are capable of processing a vast amount of data in a remarkably high speed and consider a lot of factors, called neurons, and use advanced machine learning algorithms for training. The most significant advantage comparing to other modeling methos, is that Neural Networks have the ability to be self-adaptive and transfer all the knowledge and relations among neurons during the modeling process [27].

The first appearance of Neural Networks was in 1991 by John Hertz and Anders Krogh on their book, “Introduction to the theory of Neural Computation” [28]. More specifically, each NN includes an input layer, the output layer (the prediction target) and in between the hidden layers [29]. The number of the parameters considered for each layer are translated as neurons in an ANN. The information flow regarding a specific input is related to the neurons and the weight of each one. Regarding the amount of parameters, their weights and hidden layers, Neural Networks have become a powerful tool for prediction tasks with numerous of applications and rightfully are the dominant models on the field of Artificial Intelligence [30].

## 2.4 Deep Learning

Deep Learning models are consisted of multiple processing layers to learn representations of the data [31]. The difference between a Neural Network and a Deep Neural Network lies on their structure and the complexity of the parameters as each NN with more than three hidden layers is considered as a Deep one. They are called Deep Neural Networks because thanks to the high volume of layers, and complexity the model is able to gain more important and significant information regarding the relations and patterns among the variables of the data.

The breakthrough of technology and computational power have led to the appearance of deep learning, and thus numerous Artificial Intelligence fields have been created, as well as many of the existing ones have been evolved remarkably. Since deep learning contains many hidden layers and neurons, they have become the most preferable approach to difficult, complicated problems in combination with the tremendous amount of data.

The hierarchy of this method enables the machine to learn complicated concepts by building them out of similar ones [32]. These networks use backpropagation to calculate and evaluate the weights of each neuron and especially deep convolution networks have become the most appropriate models for processing images, video, speech, and audio. The most common method for these types of predictions is the deep convolutional networks which use the notion of convolutional and pooling layers during the processing.

### 2.4.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks are a type of ANNs in that they are consisting of neurons, the input and output layer as well as the hidden layers. CNN as a deep neural network, each neuron receives an input and by performing the corresponding operations provide the desired output through a loss function [33]. CNNs are the most preferable deep learning model for Image classification, detection and segmentation problems.

Compared to ANNs, CNNs are multi-layered networks, and their most significant difference is the convolutional layers. When an image problem occurs, the model must map all the data in a mathematical perspective so as to be processed. Therefore, it translates each pixel into a value so the image will be handled as an array with its dimension to be equal to the dimension of the image. In the case of colored pictures, since each pixel corresponds to three values, one for each channel (RGB), the image will be translated into three arrays. During the training procedure, the Image data will pass through the convolutional layers in which several filters will be applied. The constant application of these filters will produce a map of activations, called feature map, and will lead to numerous transformations of the Image which will allow the CNN to identify patterns and learn to generalize its predictions [34].

It was found that Convolution Neural Networks have outstanding performance on Image classification problems. Nowadays classification has been been expanded with the ability to perform detection tasks, in which a bounding is drawn through a specific object, and even segmentation, in which a mask is predicted for each object. Their significant accuracy is due to the huge amount of labeled-annotated data that are utilized in their training procedure, such as Image Net which includes 1000 classes [35].

Consequently, CNNs have become a State-of-The-Art algorithm in computer vision. Particularly, they achieve remarkable performance on image data applications and can predict with high accuracy and efficiency image classification tasks. Regarding the objective of this thesis, CNNs are the optimal approach for the creation of an accurate Neural Network able to classify a wristwatch into a specific price class.

## 2.5 Related Work

Big Data Analytics, Machine, and Deep Learning along with their capabilities have an important effect on the operation of businesses. Particularly from an economical perspective, analytics can lead to significant insights and lead to a more profitable company. The field that focuses on financial analysis and company organizations is called Business Intelligence and Analytics (BI&A). Its key characteristic lies in the appropriate management of a company regarding the analysis of their selling data and the usage of machine learning algorithms to define effective decisions [36].

Regarding the selling price of a product, in a business manner, several attempts have been made. Approaching machine learning algorithms, companies have tried to predict the popularity of products that fits customer needs and budgets. These techniques could be Multiclass Random Forest, Multiclass Logistic Regression as well as Multiclass one-vs-all which lead to accurate and efficient predictions [37]. Although the problem of price forecasting is categorized as a regression one, these models are following the classification approach and are remarkably effective regarding the prediction of price classes and could be significantly useful in the market.

In addition, several models can lead to noteworthy predictions such as SVMs and Arima Models. These models are highly effective in cases such as house price prediction and on electronic products and they are capable of handling a tremendous amount of data in real-time compared to other machine learning algorithms [38]. Consequently, these algorithms can affect especially the field of e-commerce since it is an area that constantly produces and stores data regarding their products and their customers' preferences. A significant approach to price forecasting is Neural Networks. Artificial Neural Networks have been used in a variety of fields, including price prediction, and outcome significant results. Their accuracy is depicted in their complexity since they transfer all the information among the variables and hence the model is able to learn patterns and provide meaningful insights.

However, none of the above approaches have used these techniques on wristwatch price estimation. Companies and e-commerce in this field are constantly developing and since it concerns the majority of people independently of their annual revenue, an analysis and an AI model will be significantly useful. In addition, in most research regarding the classification approach, the true label of the class is given. This thesis not only will provide an AI classification model for price estimation but also will use unsupervised machine learning to create the corresponding classes by identifying the patterns among the watches' specifications.

# 3 Methodology

## 3.1 Dataset Description

Throughout the years countless wristwatches are being produced with constant growth in their capabilities and technologies. Therefore, the price of a typical wristwatch varies depending on its specifications, while sometimes its quality can be identified by its appearance. Nowadays, thanks to the evolution of technology and the growth of e-commerce wristwatches of all price subsets are easy to be searched and bought and hence they became accessible for all allocations.

Thanks to e-commerce, the consumer can effortlessly find a plethora of websites for ordering a wristwatch depending on his budget and desired specifications. However, the market leader especially for electronic devices and machines, is Amazon. Amazon plays a vitally important role in remarkable innovations, improvement of customer service as well as in the constant addition of new products [39]. Furthermore, it is one of the first websites that use AI technologies in order to provide more efficient recommendation systems, numerous machine and deep learning algorithms to improve its services and thus its financial revenue [40]. Amazon focuses on Artificial Intelligence, and it is not a coincidence that it is the market leader in most electronic products as well as its dominance in smart assistants with the introduction of Alexa.

The scope of this thesis is to create a Deep Neural Network and more specifically a CNN capable of categorizing a wristwatch into a price class regarding only visual data. For the efficiency of CNN, a vast amount of data is essential. Besides the visual manner of the problem, the wristwatches need to be clustered in order to be categorized in price classes. Consequently, the necessity of a dataset containing both the specifications of a wristwatch (for price considerations and clustering parameters) and its image (visual data) occurred.

The data needed for the objective of this thesis, are retrieved from Amazon and refer to wristwatches and their elements. More specifically, apart from price, a variety of variables were gathered for a sufficient amount of data for the training model and for the clustering procedure. Prior to the AI model, the watches need to be categorized into clusters by considering their characteristics. As a result, all the specifications available for each watch were gathered for a complete dataset. In addition, the image of each watch, was saved for training the CNN model. The technique for the creation of the aforesaid dataset is going to be explained extensively in the following chapter.

Apart from the image of each wristwatch all its specifications being available on Amazon's website were retrieved. For instance, for each watch variables such as its shape, its case material, its color, its band's width and thickness, case diameter, etc. are contained in the created dataset among others. More specifically, the dataset contains 4801 wristwatches with 29 variables (specifications) and they are explained extensively on Table 3 (*Appendix A*). These will contribute to the distinguishing of the preferable clusters by identifying similar patterns using unsupervised learning and the image data for the classification manner.

### 3.1.1 Data Collection and Ethical Concerns

The dataset used for the aim of this thesis, was synthesized using Amazon's data retrieved from its original website. Since there was not a publicly available dataset regarding this specific task, the necessity of the creation one occurred. The technique used for this purpose was web-scrapping, and programming language `Python` was used for its application.

Web-scrapping is a technique which allows the user to extract web content by positioning the target variables [41]. Most of the websites provide an API for the researchers to gather the desired information for a potential statistical analysis or for the training of an AI model. Nevertheless, Amazon does not provide such technology and thus the scrapping algorithm needed to be done through programming code written in python from the scratch. More specifically, the algorithm was navigated into the search pages of wristwatches. Subsequently all the URLs of each watch of the corresponding search page were retrieved until all pages finished, and were saved on a `txt` file. Finally, given the `txt` file, containing all the possible wristwatches URLs, the web-scraping algorithm retrieved and extracted all the desirable information.

Using `css` and the `html` structure of the website, the desired classes were chosen and saved on a `yml` file (which contains all the `css` for the classes). Moreover, the web-scrapping algorithm was created and with the combination of the `yml` file, the corresponding variables were gathered and saved into a JSON (`json`) file. One of the scraped elements was the image name of the corresponding watch. Its name was saved on the `json` file along with all the specification and the image was saved on a separate folder and was used on the CNN's training procedure.

The data retrieved from Amazon are publicly available on the official website. The created web-scrapping algorithm exported these data and structured them on the corresponding format for the purpose of this thesis. In addition, the images used for the creation of this dataset are also available on the Amazon's website. Consequently, no ethical concerns are occurred as all the data are publicly available and no personal information was used and their analysis aims only to academic and research purposes.

The wristwatches data, retrieved from Amazon, are 4801 with 29 variables. The data are stored in `json` format as shown on Figure 6 (*Appendix B*). This format contains a list of dictionaries, in which each dictionary represents the specifications of a wristwatch. The image of each corresponding watch was saved into JPEG (`jpg`) format on a separate folder (in which all the wristwatches images are saved) with its file name to be represented in the `json` file a key value to the dictionary. Consequently, for each observation (wristwatch) there is a dictionary with all its specification and a JPEG image, as shown on Figure 7 (*Appendix B*) saved on the corresponding folder in which all the images of the dataset are saved. The created dataset was uploaded on Kaggle and it is available upon request.

### 3.1.2 Data Preprocessing

The data exported during the web-scrapping procedure is retrieved in a raw format. In order to proceed to efficient statistical analysis and an accurate Artificial Intelligent model, significant preprocessed steps were essential.

More specifically, the majority of values of the variables were retrieved in a string format. Consequently, variables such as price, bandwidth, case thickness, item weight, etc., needed to be transformed into numeric values (integer of float). In addition, there were subjects where the unit of measurement differs on the same variable among the watches. For instance, in some circumstances, the bandwidth was measured in millimeters (mm) while in other ones in centimeters (cm). As a result, numerous modifications were performed, so each variable has the same measurement unit for all the observations.

Furthermore, there were observations in which the web-scrapping did not export the price number and thus they had a null value on the key price on the JSON file. Since the objective of this thesis is to predict the price through clustering and classification procedures, these observations were omitted. Subsequently, an id key was inserted on the dataset for each observation for convenience purposes during the analysis and the AI algorithm. The final structure of the dataset is shown in Figure 6 (*Appendix B*) and the description of each variable is depicted extensively in Table 3 (*Appendix A*).

## 3.2 Preprocessed steps

Prior to each machine learning algorithm, additional preprocessed steps are crucial for more useful insights. More specifically, an analysis of the data can lead to a better understanding of the variables and their relations and thus identify potential patterns among the specifications of wristwatches. In addition, the dataset needs to contain meaningful information for the machine and deep learning algorithms without outliers that will cause probable deviations to the models. It is correspondingly noteworthy, to exclude the null values from the dataset since it is more efficient to feed the AI models with useful information to understand the relations among the variables. Subsequently, the data frame that will be used in the following PCA and Clustering procedures will consist of 2280 observations from 6 variables (id will only be used for reference), as shown on Table 4 (*Appendix B*)

Last but not least, the clustering algorithms and their necessary preprocessed methods, such as dimensionality reduction need to measure all the possible distances among the variables to provide meaningful insights. Consequently, a normalization of the data prior to every algorithm should be performed. Normalization is a procedure in which the data are reorganized in a way that there is no redundancy, and all the related data are stored together [42]. Normalization also leads to the avoidance of bias and contributes to admiringly accurate results by providing less computational power. This process performs on every distanced-based method efficiently and hence it will be used prior to clustering and Principal Component Analysis to deliver more efficient and objective results.

## 3.3 Clustering Procedure

The aim of this thesis is to create an Artificial Intelligence model in order to predict the price class of a wristwatch accurately. However, in the created dataset these classes are not predefined. For that purpose, unsupervised machine learning technique is going to be utilized to categorize the watches into clusters. Afterwards, these clusters will be considered as classes and as the ground truth for the corresponding AI model.

The purpose of the categorization is to estimate the price of a wristwatch depending on the class that it belongs. Consequently, the variable price plays a vital role for the aforesaid clustering. In addition, a variety of specifications that affects the price, are also going to be used to lead to the optimal result of the clustering model. These are the Case Diameter, Case Thickness, Band Width, Item Weight and Water Resistant depth.

Since the number of the desired clusters is not defined, hierarchical clustering will be used as the unsupervised technique due to its better efficiency on unknown categories in comparison to k-means algorithms. Hierarchical Clustering and all clustering approaches need to measure all the possible distances among the variables. For the thesis task, six variables are provided for defining the number clusters. Thus, the problem is translated as six dimensional one which requires high computational power and calculation complexity for the computer. Therefore, a dimensionality reduction method facilitates to a less complex problem by reducing the number of the variables without losing meaningful information.

### 3.3.1 Principal Component Analysis (PCA)

As it was extensively described on the Literature Review, Principal Component Analysis is a preprocess step that reduces the dimensionality of the problem by keeping the desired percent of information. The normalization of the data has been preceded and thus the PCA can be performed.

Performing the PCA algorithm into the variables regarding the price of a wristwatch and its specifications that affect it, the eigen values were calculated. More specifically, the considered variables for the PCA method were the Price, Case Diameter, Case Thickness, Band Width, Item Weight and Water-Resistant depth (six in total). It was concluded that choosing 4 features, the 99% of the information is kept. Consequently, four Principal Components needed to be created to preserve the meaningful information. Based on Figure 1. it is observed that approximately 90% of the information is also kept with 3 Principal Components. However, to achieve the optimal results, the thesis will follow the most accurate approach and thus the 4 PCs were chosen.

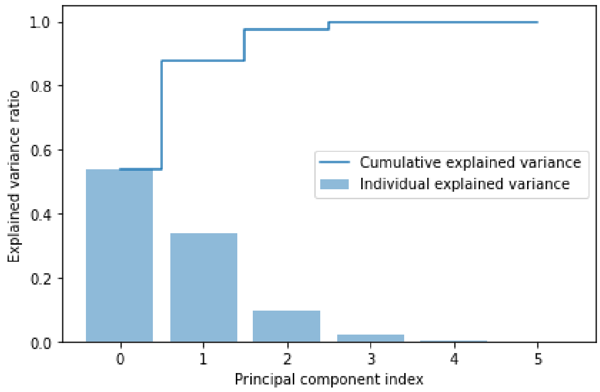


Figure 1: Principal Components - percent of the information kept

Therefore, four new variables (Principal Components) were created, which represents the aforesaid six ones. These new components will be used both for the further analysis and the clustering procedure. A random selection of the table of the principal components along with the id of each product is displayed on Table 5 (*Appendix C*).

### 3.3.2 Hierarchical Clustering

Hierarchical Clustering (HC) is an unsupervised learning technique that categorizes the variables into clusters by a distance-based algorithm using Agglomerative method. Its significant advantage, compared to other methods, lies in the non-necessary definition of the number of clusters. HC is highly affected by the normalization of the data and the combination of the prior Principal Component Analysis will provide the optimal results clusters. In addition, Ward's linkage will be selected. This choice is considered for hierarchical clustering as the most suitable since it relies on the variance of data to keep the merge distance as close to zero as possible, despite its constant increasing.

The corresponding hierarchical tree will not consider the initial 6 variables for its structure whereas the 4 Principal Components provided with PCA are going to be utilized. More specifically, these components contain more than 99% of the information and in fact, have reduced the dimensionality of the problem to two. This approach will contribute to more efficient results and will reduce the complexity of the problem as well as the computational power. The constructed tree using 2280 observations from 4 variables (Principal Components) is shown in Figure 8 (*Appendix D*). The x-axis of the dendrogram portrays the variables with the dendrites to be the possible clusters while the y-axis refers to the distance of centers from all the observations. Considering the distance measured, on the depicted tree, it is significant to prune the tree into nine clusters (dendrites) which correspond to a value of y approximately 10000 as depicted in Figure 2 with a red line.

Pruning the tree into nine clusters means that watches that belong to the same cluster will most likely have similar characteristics. For instance, the scatterplot of the Price and Item Weight variables, produced after the clustering categorization, is shown in Figure 9 (*Appendix D*). The average price for

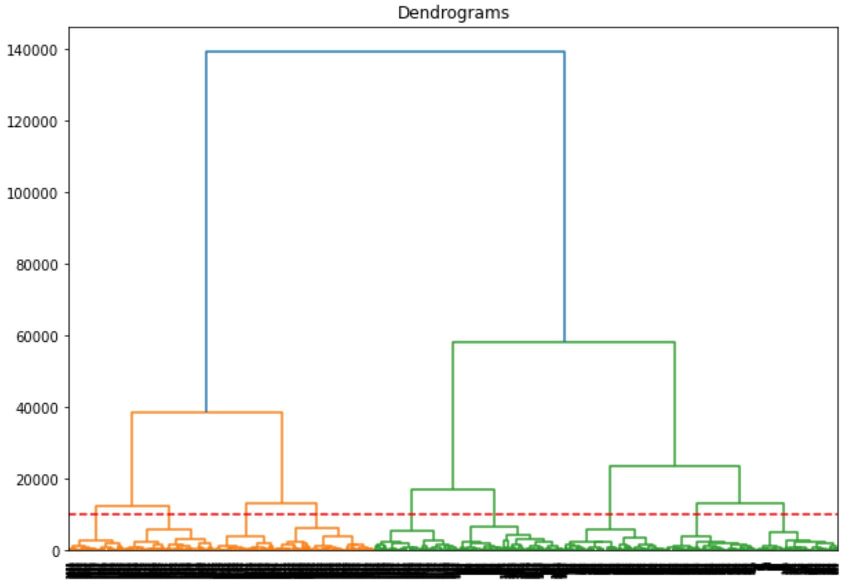


Figure 2: Dendrogram of Principal Components - Pruning Line

each wristwatch is depicted on Table 7 (*Appendix D*).

Due to the efficiency of the clustering, the wristwatches can be placed into the appropriate classes with each observation (wristwatch) to be categorized based on its depicted cluster. Since there are nine clusters produced via machine learning, the classes for the deep learning model will also be nine. Depending on the corresponding class-cluster the image name of each watch is retrieved. Then, through each filename, the watches are also categorized through the visual data. More specifically, after the Hierarchical Clustering, a python file was created to parse automatically the visual data into a number of folders equal to the number of clusters. Consequently, the images were parsed into nine folders in which each folder was named with the name of the class (such as 0, 1,..,8) and contained the corresponding images (the ones that belong to the specific class). The number of watches that each class contain is shown in Table 2. For instance, a folder named 0 was created, which contains 319 images, the ones that belong to class 0. This strategy was also applied to the remaining classes

| Class | Watches |
|---|---|
| 0 | 319 |
| 1 | 381 |
| 2 | 245 |
| 3 | 284 |
| 4 | 199 |
| 5 | 178 |
| 6 | 189 |
| 7 | 207 |
| 8 | 149 |

Table 1: Number of watches per class

Considering the above method, the clustering leads to the definition of price classes for the objective of this thesis. Each data which illustrates a wristwatch besides its characteristics, its image belongs to a certain category (class), as demonstrated on Table 6. (*Appendix D*). Consequently, since the visual data have been categorized the CNN model have the required information (image data along with their ground truth class) for its training.

## 3.4 Convolutional Neural Network

Prior to the training of the model, the split of dataset into training and test sets is essential. The training set will be consisted of the 70% of the data while the test set the remaining 30% one. The purpose of this model is to predict the price class that was provided through Hierarchical Clustering given as input image data. Consequently, the corresponding data are visual data that are referring to the 2280 observation from the wristwatches' specifications. For the training procedure, 1507 images will be used and the remaining data will test the efficiency of the created model.

As it was described on the preprocessed steps chapter, the image data are organized into 9 folders regarding their belonging class. Therefore, the training data will be consisted of the 70% of the data per each folder – class. Two folders are going to be created one for the training and one for the test set respectively. Each folder will contain 9 folders with the corresponding number of images, to distinguish their belonging class. The representation of the amount of images per set and class is depicted on Table 2.

| Class | Training Set | Test Set |
|---|---|---|
| 0 | 223 | 96 |
| 1 | 266 | 115 |
| 2 | 172 | 73 |
| 3 | 200 | 84 |
| 4 | 140 | 59 |
| 5 | 125 | 53 |
| 6 | 132 | 57 |
| 7 | 145 | 62 |
| 8 | 104 | 45 |
| **Sum** | **1507** | **644** |

Table 2: Number of images for training and test sets

Since the model will be trained regarding RGB images, the visual data will be translated into three arrays in which each pixel is representing each value of the Red, Green, and Blue channels. The model will parse these data into complex augmentations and through convolutional and pooling layers, a Deep Neural Network (and specifically Convolutional Neural Network) will identify patterns and be able to generalize on unlabeled data. Thus, for the training procedure since the images are handled as arrays, they need to be on the same dimension. All images, from the created dataset, are of dimension (445,266,3). The last one represents the number of color channels indicating that they are RGB images.

Using `Python`, the most preferable libraries for Deep Learning algorithms are `TensorFlow` and `PyTorch`. For this thesis approach, the Convolutional Neural Network will be written using `PyTorch`. Although `TensorFlow` is more comprehensive for inexperienced and less complicated compared to `PyTorch`, the last one provides more insights by giving the researcher access to all the layers and the neurons of the network leading to optimal results.

An essential step in `PyTorch`, prior to the definition of the network is the creation of a class that represents the data, called the data loader [43]. Using this class, the model identifies the path of the training and test set, their structure, the parsing among the variables, and their ground truth (price classes). In addition, it is highly important to define the method of reading the input value, since image data first need to be translated into arrays (using the `OpenCV` library). Subsequently, a variety of transformations have been performed before the structure of the CNN. Since there is no difference in the dimension among the images, they do not need to be resized. However, the images were resized to the dimension of 292x171 (approximately half of their original size), as shown on Figure 3. to reduce the complexity of the model and increase its training speed. In addition, a normalization was also essential for a more efficient training of the neural network.

Afterward, the Convolutional Neural Network was defined. The data will pass through two convolutional layers, with a pooling layer between them, and finally on three fully connected layers. In a convolution layer, a kernel is shifted along the input matrix which using the dot product between the vectors, it outputs a scalar and thus it results in a different dimension array maintaining the information of the image [44]. On the other hand, pooling layers downsampling the size of the input image (and


Figure 3: Image Sizes for the Training Procedure (292,171)

more specifically its created feature map) leads to the reduction of the network's computations. There are several methods of pooling, with the most common and the one that this thesis used, is Max Pooling. Max pooling uses the largest pixel value among all the pixels from the input image [45]. Finally, fully connected layers take the output of the previous layers and flatten it into a single vector used in the upcoming stage, and thus it depicts the features of the image in only a one-dimensional array [46].

Moreover, the loss function as well as the optimizer needed to be specified. For the loss function, Cross Entropy was selected which measures the loss of the predicted value to its ground truth. In the case of images where arrays encounter, cross-entropy estimates the probability distribution between the variables [36]. Last but not least, the chosen optimizer was Stochastic Gradient Descent (SGD) which is recommended for classification problems as it descends the slope to reach the lowest point on the observation's graphs [37]. SGD describes the loss of the model regarding the ground truth, therefore the purpose of the model is to minimize it. Subsequently, from a mathematical perspective, this procedure translates into finding the minimum point so the one in which the derivative equals zero. Consequently, it starts calculating the derivative on each point step by step. The number that describes how the size of the aforesaid step changes, is called the learning rate. For this Stochastic Gradient Descent, the chosen learning rate was 0.01.

While the network is processing image data, it is more preferable to be trained on the graphics cars (GPU) rather than the CPU while it reduces the training time remarkably. As a result, the proposed model was trained on GPU, and especially with Cuda 11.6 .

Taking all the aforesaid parameters into account, the model was trained using a batch size of 4 images in 25 epochs on the 'Cuda' device. This batch size indicates that in each epoch, 4 images pass towards the network each time so this procedure is repeated 395 times to use all training images (1507). The training technique along with the modification of loss in each epoch is depicted in Figure 10(*Appendix E*). The training procedure needed approximately 11 minutes and the corresponding model succeeded to reduce the loss to 0.0369. To ensure the efficiency of training, validation data were also used and thus both training and validation losses were constantly measured. In cases of overfitting, although the training loss is constantly reducing, the validation loss up to a certain point starts to increase indicating that the model learns accurately and explicitly the training data without the possibility of generalization. Based on Figure 4 it is concluded that the training was objective and it did not lead to overfitting since both lines were reducing their errors.

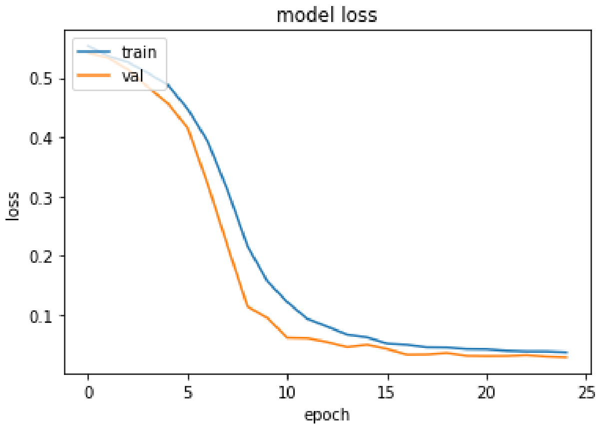


Figure 4: Training and Validation Losses

# 4 Analysis and Discussion

The aforesaid Convolution Neural Network was created to predict the price class of a new unlabeled wristwatch by using only its front view image. Using the test set, which consists of 30% of the data (773 images) as input to the developed model, it is concluded that achieves 93% accuracy. More precisely, the accuracy of each class is depicted in Table 8 (*Appendix F*) . Consequently, the model can predict efficiently and highly accurately the price class of a new wristwatch. For instance, as shown in Figure 5, 4 random wristwatches are considered along with their ground truth belonging class. Considering these images, the model predicted the classes "1 2 2 2" (on the corresponding order) which are precisely identical to the ground truth.


Figure 5: Sample of 4 wristwatches along with their corresponding classes

However, for the identification of optimal results, several modifications were performed to test their effects on the training procedure. More specifically, in the Principal Component Analysis, it was indicated that except for the 4 principal components (PCs), the chosen 3 ones could also preserve valuable information. Using 3 principal components, 93% of the meaningful information remains for processing. Since PCA reduces the dimension of the problem, and 93% reaches a satisfying percentage of the information, the selection of 3 components could also deliver objective results while it reduces the complexity of the problem. Considering this parametrization the training procedure was again performed. The created CNN using the above approach as well as keeping the same optimizer, loss function, batch size, and the number of epochs, led to 90% accuracy, which is also an admiringly effective model. However, their difference of 3% in the accuracy, may lead to probable miscategorization of data, thus the first model will be chosen as the most preferable for the objective of this thesis.

# 5 Conclusion

Artificial Intelligence nowadays has become one of the most rapid evolution fields that have spurred the interest of many researchers. Their applications resulting in accurate predictions have affected vitally the modern society. Thanks to the remarkable growth of technology, computers, and machines have exceeded their potentials and thus can perform complex mathematical procedures at a significant speed and thus enhancing the field of Artificial Intelligence.

Machine and Deep Learning are the most common approaches to Artificial Intelligence. Their main scope is accurate predictions by considering a significant amount of data, calculating their relations, and identifying probable patterns among them. Consequently, during their training process, the number of data plays a vital role in the efficiency of the model. Thanks to Big Data, a tremendous amount of data is now publicly available through the constant gathering of information for various devices, websites, applications, etc. As a result, more data can be fed into the training procedure, and with the combination of the computers evolution in the way they process data, and the increase of the considerable batch sizes, the accuracy of the AI models has been remarkably improved.

The most applicable forecasting technique with plenty of capabilities, is Neural Network. NNs can handle complex information, identify patterns and pass all the required knowledge through neurons. Moreover, complex Neural Networks with more than three hidden layers (Deep Neural Networks) such as Convolutional Neural Networks dominate the field of computer vision. Particularly, CNNs due to convolutional and pooling layers, perform with notable accuracy in classification, detection, and segmentation tasks referring to image data.

Besides the extended research application of Artificial Intelligence, forecasting algorithms can also be valuable in an industrial manner. Machine and Deep Learning can be applied to forecast the annual revenue of a company. Apart from analytics, companies can use AI models for reaching their utmost income regarding the optimal management of their productivity as well as the effect of potential decisions that may affect their revenue. This thesis provided an accurate Deep Learning model in an industry manner regarding the price estimation of a wristwatch. Wristwatches are daily used machines that play a vital role in organizing the daily routine. Since they are highly useable, their price range has been significantly exceeded, so all individuals can procure their preferable watch depending on its specifications.

A dataset of various watches was created with their specifications along with their images, using products exported from Amazon, which is the market leader in the field of e-commerce. Utilizing these data, price classes were created automatically using unsupervised machine learning, by identifying the patterns among the watches' specifications. Furthermore, several preprocessed steps were conducted to reduce the complexity of the problem, such as Principal Component Analysis. Thereafter, a Convolutional Neural Network was trained using the data and the aforesaid classes as the ground truth to generalize on new unlabeled data. The created model was highly efficient as it can classify a new watch with 93

Considering the above procedure, this thesis proposed a classification approach for a price estimation of a wristwatch. The average price depending on the predicted class, is depicted on Table 7 (*Appendix D*) This model can be applied by companies to achieve the desired design of a new watch that leads to its utmost selling price. For instance, in case the corresponding price is not satisfactory, the product can be redesigned and fed again to the model and by continuing this approach, the most optimal design will be achieved.

# 6 Future Steps

The objective of this thesis was to create an accurate forecasting price estimation model using only visual data. The created Convolutional Neural Network as well as the variety of preprocessed steps and prior analysis led to high accuracy. The model can predict the price class of a new wristwatch considering only its front view with 93% accuracy. Consequently, it could be significantly beneficial for companies focusing on the domain of wristwatches. Its application lies in the fact of prior knowledge of the price by only using a computer-designed image prior to the productivity level.

Although the model succeeded high accuracy, it uses only visual data as input. Despite its efficiency, the price of a wristwatch is depending on a variety of factors and specifications such as color, size, smart possibilities, etc. While these specifications are considered for the definition of clusters (which are then used as classes) they are not given as input to the created model. Consequently, a future approach in which the model could receive multiple inputs may lead to more accurate results. For instance, the model could consider these specifications and handle them as a multi-classification problem and with max voting, it could define the predicted class and hence the price estimation.

On the other hand, since the goal is price estimation, another approach for defining the price of a new wristwatch could be the regression one. Regression problems in contrast to classification ones are able to predict an exact value of the desired output, such as an exact estimation of the price. Consequently, a future step of the thesis objective is the regression approach which will provide more meaningful insights into the selling price as the model will predict a number that represents the potential price as accurately as possible and not a class that has a range of values.

Last but not least, the presented approach of this thesis could be beneficial in an industrial manner regarding companies annual revenue. More specifically, this model could be generalized to more products and not restrictively on wristwatches, expanding the field of its applications. For instance, it could be trained on the whole list of an e-commerce website's products and thus can offer extensive usefulness to the corresponding company by providing a more structured and generalized AI price prediction model. Finally, an interface could provide a more user-friendly environment in which the manager, simply by importing an image, will get a selling price forecasting along with its percentage of accuracy. All the above will be highly efficient on an industrial level, mainly for financial benefits, and will be left as a future step for their implementation.

## References


[1] M. Bellis, “Who invented the clock?” Feb 2019. [Online]. Available: https://www.thoughtco.com/clock-and-calendar-history-1991475

[2] V. Niculescu, “On the impact of high performance computing in big data analytics for medicine,” *Applied Medical Informatics*, vol. 42, pp. 9–18, 05 2020.

[3] W. Tian and Y. Zhao, “2 - big data technologies and cloud computing,” in *Optimized Cloud Resource Management and Scheduling*, W. Tian and Y. Zhao, Eds. Boston: Morgan Kaufmann, 2015, pp. 17–49. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128014769000021

[4] A. Williams, “Why are rolex watches even more expensive right now?” Mar 2022. [Online]. Available: https://www.nytimes.com/2022/03/24/style/rolex-watches-cost-boom.html?smid=url-share

[5] S. Sagiroglu and D. Sinanc, “Big data: A review,” in *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 2013, pp. 42–47.

[6] I. Yaqoob, I. A. T. Hashem, A. Gani, S. Mokhtar, E. Ahmed, N. B. Anuar, and A. V. Vasilakos, “Big data: From beginning to future,” *International Journal of Information Management*, vol. 36, no. 6, Part B, pp. 1231–1247, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401216304753

[7] yourtechdietAdmin, “7 v’s of big data explained (along with infographic),” Nov 2021. [Online]. Available: https://yourtechdiet.com/blogs/7vs-big-data/#Variability

[8] “What is artificial intelligence (AI)? definition, benefits and use cases.” [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

[9] J. Redmon. [Online]. Available: https://pjreddie.com/darknet/yolo/

[10] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[11] M. Zimmer, P. Viappiani, and P. Weng, “Teacher-student framework: A reinforcement learning approach,” 05 2014.

[12] J. M. Carew, “What is reinforcement learning? a comprehensive overview,” Mar 2021. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning

[13] “What is machine learning and why is it important?: Micro focus.” [Online]. Available: https://www.microfocus.com/en-us/what-is/machine-learning

[14] E. Burns, “What is machine learning and why is it important?” Mar 2021. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML

[15] I. H. Sarker, "Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems," *SN Computer Science*, vol. 3, no. 2, 2022.

[16] M. D. Schluchter, *Mean Square Error.* John Wiley, Ltd, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011815.b2a15087

[17] B. I. C. Education, "What is supervised learning?" [Online]. Available: https://www.ibm.com/cloud/learn/supervised-learning

[18] P. Cunningham, M. Cord, and S. J. Delany, *Supervised Learning.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 21–49. [Online]. Available: https://doi.org/10.1007/978-3-540-75171-7_2

[19] B. I. C. Education, "What is unsupervised learning?" [Online]. Available: https://www.ibm.com/cloud/learn/unsupervised-learning

[20] Z. Ghahramani, *Unsupervised Learning.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 72–112. [Online]. Available: https://doi.org/10.1007/978-3-540-28650-9_5

[21] K. P. Sinaga and M.-S. Yang, "Unsupervised k-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80 716–80 727, 2020.

[22] F. Nielsen, *Hierarchical Clustering.* Cham: Springer International Publishing, 2016, pp. 195–211. [Online]. Available: https://doi.org/10.1007/978-3-319-21903-5_8

[23] Vijaya, S. Sharma, and N. Batra, "Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 568–573.

[24] "Agglomertive hierarchical clustering using ward linkage - github pages." [Online]. Available: https://jbhender.github.io/Stats506/F18/GP/Group10.html

[25] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.101

[26] I. Jolliffe, *Principal Component Analysis.* John Wiley & Sons, Ltd, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa501

[27] A. Jain, J. Mao, and K. Mohiuddin, "Artificial neural networks: a tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996.

[28] J. A. Anderson, *An introduction to neural networks.* MIT Press, 1995.

[29] [Online]. Available: https://www.sas.com/el_gr/insights/analytics/neural-networks.html

[30] C. M. Bishop, "Neural networks and their applications," *Review of Scientific Instruments*, vol. 65, no. 6, pp. 1803–1832, 1994. [Online]. Available: https://doi.org/10.1063/1.1144830

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," May 2015.

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

[33] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015. [Online]. Available: https://arxiv.org/abs/1511.08458

[34] J. Brownlee, "How do convolutional layers work in deep learning neural networks?" Apr 2020. [Online]. Available: https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/

[35] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.

[36] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012. [Online]. Available: http://www.jstor.org/stable/41703503

[37] R. I. Naeem Ahmed Mahoto, "An intelligent business model for product price prediction using machine learning approach," *Intelligent Automation & Soft Computing*, vol. 30, no. 1, pp. 147–159, 2021. [Online]. Available: http://www.techscience.com/iasc/v30n1/43971

[38] K.-K. Tseng, R. F.-Y. Lin, H. Zhou, K. J. Kurniajaya, and Q. Li, "Price prediction of e-commerce products through Internet sentiment analysis," *Electronic Commerce Research*, vol. 18, no. 1, pp. 65–88, March 2018. [Online]. Available: https://ideas.repec.org/a/spr/elcore/v18y2018i1d10.1007_s10660-017-9272-9.html

[39] B. Singh, P. Kumar, N. Sharma, and K. P. Sharma, "Sales forecast for amazon sales with time series modeling," in *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, 2020, pp. 38–43.

[40] B. Smith and G. Linden, "Two decades of recommender systems at amazon.com," *IEEE Internet Computing*, vol. 21, no. 3, pp. 12–18, 2017.

[41] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, and F. Fdez-Riverola, "Web scraping technologies in an API world," *Briefings in Bioinformatics*, vol. 15, no. 5, pp. 788–797, 04 2013. [Online]. Available: https://doi.org/10.1093/bib/bbt026

[42] Techopedia, Aug 2020. [Online]. Available: https://www.techopedia.com/definition/1221/normalization

[43] “Datasets & dataloaders.” [Online]. Available: https://pytorch.org/tutorials/beginner/basics/data_tutorial.html

[44] D. Unzueta, Mar 2022. [Online]. Available: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

[45] V. Lakkavaram, L. Raghuveer, C. Kumar, G. Sri, and S. Habeeb, “A review on practical diagnostic of tomato plant diseases,” 04 2019.

[46] “Fully connected layers in convolutional neural networks,” Jun 2022. [Online]. Available: https://indiantechwarrior.com/fully-connected-layers-in-convolutional-neural-networks/

# Appendix A Description of Dataset

The dataset of Table 3 consists of 4,803 records with 29 variables regarding wristwatches retrieved from Amazon's website.

| | Variable Name | Type | Description |
|---|---|---|---|
| 1. | Id | Integer | The id of each observation |
| 2. | Name | String | The name of each wristwatch as given on amazon's website |
| 3. | Image Name | String | The exact name of the image of the watch, including its format |
| 4. | Price | Float | The price of wristwatch |
| 5. | Brand, Seller or Collection Name | String | The brand name of the corresponding wristwatch |
| 6. | Model Number | String | The serial number of the wristwatch |
| 7. | Part Number | String | The id number of the subcategory model |
| 8. | Item Shape | String | Describes the shape of the wristwatch (e.g., Round) |
| 9. | Dial Window Material Type | String | The type of watch's window material |
| 10. | Display Type | String | Describes the type of display (e.g., analog) |
| 11. | Clasp | String | The type of Clasp (e.g., Jewelry Clasp) |
| 12. | Case Material | String | Describes the case material |
| 13. | Case Diameter | Float | The diameter of the wristwatch's case in millimeters (mm) |
| 14. | Case Thickness | Float | The thickness of the wristwatch's case in millimeters (mm) |
| 15. | Band Material | String | The band material (e.g., steel) |
| 16. | Band Size | String | Categorizes the band size (e.g., Men's standard) in meters (m) |

| | | | |
|---|---|---|---|
| 17. | Band width | Float | The band's width in millimeters (mm). |
| 18. | Band Color | String | Describes the band's color (e.g., gold) |
| 19. | Dial Color | String | Describes the dial's color (e.g., black) |
| 20. | Bezel Material | String | Describes the bezel's material |
| 21. | Bezel Function | String | Describes the possible functions (e.g., 12-hour time display) |
| 22. | Calendar | String | The type of calendar (e.g., Day) |
| 23. | Item Weight | Float | The wristwatch's weight in grams (gr) |
| 24. | Movement | String | The type of movement (e.g., Quartz) |
| 25. | Water Resistant Depth | Float | The maximum depth of water in meters (m) |
| 26. | ASIN | String | The Amazon Serial Identification Number |
| 27. | GPS | String | The type of the provided GPS. No in case there is not GPS. |
| 28. | Manufacturer | String | The name of the manufacturer |
| 29. | Colour | String | The wristwatch's color |

Table 3: Description of Variables

```
{
    "id": 1,
    "Name": "Casio Men's F105W-1A Illuminator Sport Watch",
    "Image Name": "51GUx+NFHgL._AC_UY445_.jpg",
    "Price": 19.73,
    "Brand, Seller, or Collection Name": "Casio",
    "Model Number": "EAW-F-105W-1A",
    "Part Number": "EAW-F-105W-1A",
    "Item Shape": "Square",
    "Dial Window Material Type": "Mineral",
    "Display Type": "Digital",
    "Clasp": "Buckle",
    "Case Material": "Resin",
    "Case Diameter": 35.0,
    "Case Thickness": 10.0,
    "Band Material": "Plastic",
    "Band Size": "Mens-Standard",
    "Band Widht": 18.0,
    "Band Color": "Black",
    "Dial Color": "Digital",
    "Bezel Material": "Resin",
    "Bezel Function": "Stationary",
    "Calendar": "Day-Date",
    "Item Weight": 21.0,
    "Movement": "Quartz",
    "Water Resistant Depth": 50,
    "ASIN": "B000GB1RAU",
    "GPS": "No",
    "Manufacturer": "Casio",
    "Colour": "Black"
},
{
    "id": 2,
    "Name": "Fossil Gen 5 Carlyle Stainless Steel Touchscreen Smartwatch with Speaker, Heart Rate, GPS, Contactless Payments, and Smartphone Notifications",
    "Image Name": "71tc0IQKJGL._AC_UX522_.jpg",
    "Price": 196.62,
    "Brand, Seller, or Collection Name": "Fossil",
    "Model Number": "FTW4026",
    "Part Number": "FTW4026",
    "Item Shape": "Round",
    "Dial Window Material Type": "Synthetic sapphire",
    "Display Type": "Analog",
    "Clasp": "Fold-Over Clasp with Double Push-Button Safety",
    "Case Material": "Stainless Steel",
    "Case Diameter": 44.0,
    "Case Thickness": 12.0,
    "Band Material": "Leather",
    "Band Size": "Mens-Standard",
    "Band Widht": 22.0,
    "Band Color": "Brown",
    "Dial Color": "Blue",
    "Bezel Material": "Steel",
    "Bezel Function": "Stationary",
    "Calendar": "Special Feature",
    "Item Weight": 79.94565514199999,
    "Movement": "Touchscreen",
    "Water Resistant Depth": 30.0,
    "ASIN": "B07VWL5V1T",
    "GPS": "Special Feature",
    "Manufacturer": "Fossil",
    "Colour": "Brown"
},
```


Figure 6: Structure of Dataset


```
{
    "id": 2,
    "Name": "Fossil Gen 5 Carlyle Stainless Steel Touchscreen Smartwatch with Speaker, Heart Rate, GPS, Contactless Payments, and Smartphone Notifications",
    "Image Name": "71tc0IQKJGL._AC_UX522_.jpg",
    "Price": 196.62,
    "Brand, Seller, or Collection Name": "Fossil",
    "Model Number": "FTW4026",
    "Part Number": "FTW4026",
    "Item Shape": "Round",
    "Dial Window Material Type": "Synthetic sapphire",
    "Display Type": "Analog",
    "Clasp": "Fold-Over Clasp with Double Push-Button Safety",
    "Case Material": "Stainless Steel",
    "Case Diameter": 44.0,
    "Case Thickness": 12.0,
    "Band Material": "Leather",
    "Band Size": "Mens-Standard",
    "Band Widht": 22.0,
    "Band Color": "Brown",
    "Dial Color": "Blue",
    "Bezel Material": "Steel",
    "Bezel Function": "Stationary",
    "Calendar": "Special Feature",
    "Item Weight": 79.94565514199999,
    "Movement": "Touchscreen",
    "Water Resistant Depth": 30.0,
    "ASIN": "B07VWL5V1T",
    "GPS": "Special Feature",
    "Manufacturer": "Fossil",
    "Colour": "Brown"
},
```



71tc0IQKJGL._AC_UX522_.jpg

Figure 7: Representation of an observation along with its corresponding image

## Appendix B Statistical Analysis and Preprocessed Steps

| id | Price | Case Diameter | Case Thickness | Band Width | Item Weight | Water Res. Depth |
|---|---|---|---|---|---|---|
| 4 | 101.38 | 32.0 | 7.0 | 14.0 | 98.939836 | 50.292 |
| 5 | 49.90 | 40.0 | 12.0 | 20.0 | 250.042794 | 201.168 |
| 6 | 149.27 | 20.5 | 7.1 | 17.0 | 89.867988 | 30.0 |
| 7 | 35.00 | 32.0 | 8.5 | 16.0 | 54.431084 | 50.0 |
| 8 | 219.00 | 32.0 | 10.1 | 18.0 | 725.747792 | 100.0 |
| ... | ... | ... | ... | ... | ... | ... |
| 4795 | 130.21 | 31.9 | 6.8 | 16.9 | 28.349523 | 30.48 |
| 4798 | 144.99 | 37.0 | 14.0 | 18.0 | 102.058283 | 201.168 |
| 4800 | 86.09 | 53.0 | 16.7 | 25.8 | 7.0873808 | 201.168 |
| 4802 | 305.00 | 30.0 | 6.5 | 13.0 | 59.699046 | 15.24 |
| 4803 | 186.28 | 33.0 | 11.0 | 10.0 | 100.073817 | 100.584 |

Table 4: Omitted null values - 2208 observations

## Appendix C Principal Component Analysis

| id | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| 4 | -3246.888925 | -117.639793 | 11.726355 | -41.838526 |
| 5 | -3246.781623 | -33.567459 | 155.173215 | 102.083549 |
| 6 | -3244.256743 | -92.042482 | -28.122453 | -65.689444 |
| 7 | -3241.890669 | -195.744140 | 26.313771 | -33.904924 |
| 8 | -3245.755345 | 407.222671 | 377.421065 | -44.562766 |
| ... | ... | ... | ... | ... |

Table 5: Principal Components along with the id

## Appendix D Clustering Procedure

| id | Price | Case Diam. | Case Thick. | Band Width | Item Weight | Water Depth | Class |
|---|---|---|---|---|---|---|---|
| 4 | 101.38 | 32.0 | 7.0 | 14.0 | 98.939836 | 50.292 | 0 |
| 5 | 49.90 | 40.0 | 12.0 | 20.0 | 250.042794 | 201.168 | 0 |
| 6 | 149.27 | 20.5 | 7.1 | 17.0 | 89.867988 | 30.0 | 0 |
| 7 | 35.00 | 32.0 | 8.5 | 16.0 | 54.431084 | 50.0 | 0 |
| 8 | 219.00 | 32.0 | 10.1 | 18.0 | 725.747792 | 100.0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 4795 | 130.21 | 31.9 | 6.8 | 16.9 | 28.349523 | 30.48 | 2 |
| 4798 | 144.99 | 37.0 | 14.0 | 18.0 | 102.058283 | 201.168 | 2 |
| 4800 | 86.09 | 53.0 | 16.7 | 25.8 | 7.0873808 | 201.168 | 2 |
| 4802 | 305.00 | 30.0 | 6.5 | 13.0 | 59.699046 | 15.24 | 2 |
| 4803 | 186.28 | 33.0 | 11.0 | 10.0 | 100.073817 | 100.584 | 2 |

Table 6: Wristwatch specifications along with the corresponding classes

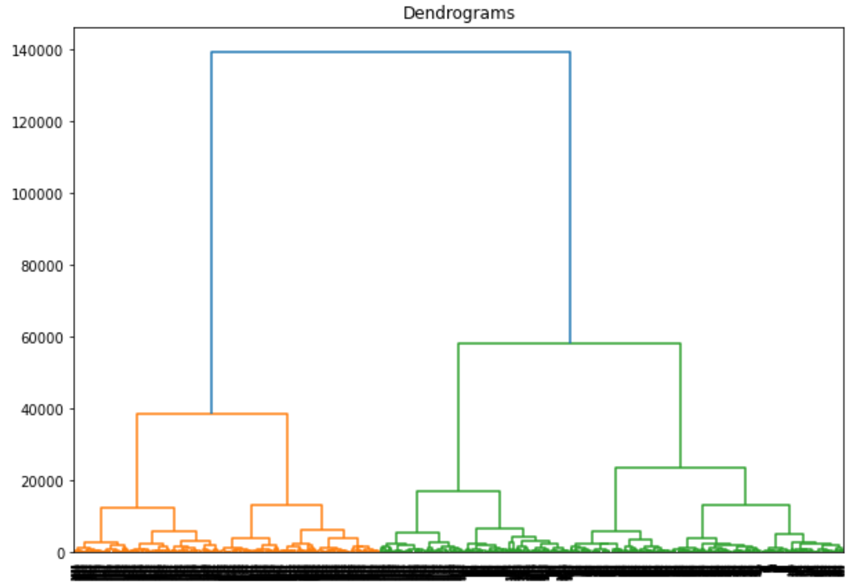


Figure 8: Dendrogram using Principal Components

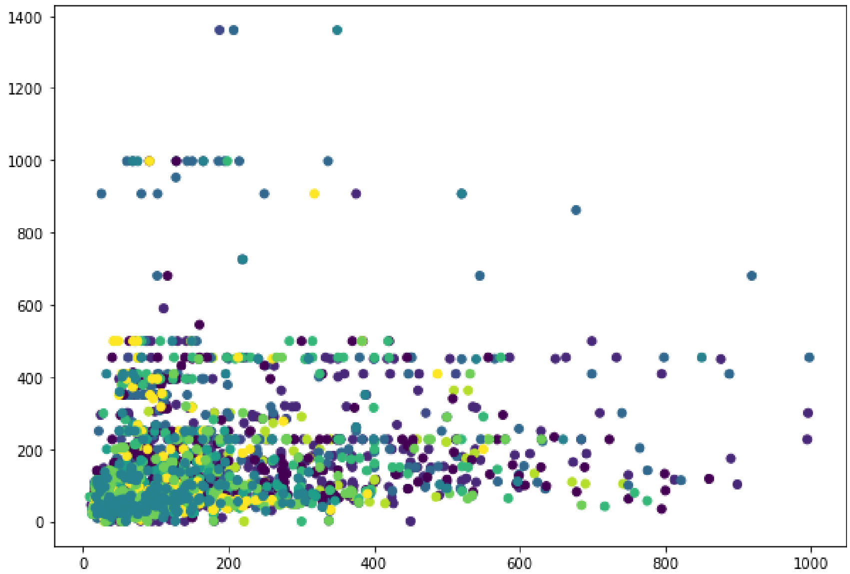


Figure 9: Scatterplot of Price and Item Weight

| Class | Average Price ($) |
|---|---|
| 0 | 129.415 |
| 1 | 207.377 |
| 2 | 166.854 |
| 3 | 127.684 |
| 4 | 204.024 |
| 5 | 130.711 |
| 6 | 188.761 |
| 7 | 113.669 |
| 8 | 194.167 |

Table 7: Wristwatch's average price per class

# Appendix E Convolution Neural Network

```
cuda:0

Epoch: 1/25: 100%|██████████| 395/395 [00:33<00:00, 11.67batch/s, loss=0.553]
Epoch: 2/25: 100%|██████████| 395/395 [00:20<00:00, 19.47batch/s, loss=0.536]
Epoch: 3/25: 100%|██████████| 395/395 [00:18<00:00, 21.65batch/s, loss=0.526]
Epoch: 4/25: 100%|██████████| 395/395 [00:19<00:00, 20.31batch/s, loss=0.508]
Epoch: 5/25: 100%|██████████| 395/395 [00:18<00:00, 21.58batch/s, loss=0.487]
Epoch: 6/25: 100%|██████████| 395/395 [00:20<00:00, 19.73batch/s, loss=0.447]
Epoch: 7/25: 100%|██████████| 395/395 [00:18<00:00, 21.77batch/s, loss=0.393]
Epoch: 8/25: 100%|██████████| 395/395 [00:20<00:00, 19.35batch/s, loss=0.31]
Epoch: 9/25: 100%|██████████| 395/395 [00:18<00:00, 21.59batch/s, loss=0.216]
Epoch: 10/25: 100%|██████████| 395/395 [00:17<00:00, 22.79batch/s, loss=0.156]
Epoch: 11/25: 100%|██████████| 395/395 [00:17<00:00, 22.41batch/s, loss=0.122]
Epoch: 12/25: 100%|██████████| 395/395 [00:20<00:00, 18.89batch/s, loss=0.0933]
Epoch: 13/25: 100%|██████████| 395/395 [00:19<00:00, 20.48batch/s, loss=0.0803]
Epoch: 14/25: 100%|██████████| 395/395 [00:19<00:00, 20.53batch/s, loss=0.0667]
Epoch: 15/25: 100%|██████████| 395/395 [00:19<00:00, 20.77batch/s, loss=0.0621]
Epoch: 16/25: 100%|██████████| 395/395 [00:18<00:00, 21.52batch/s, loss=0.0518]
Epoch: 17/25: 100%|██████████| 395/395 [00:20<00:00, 19.33batch/s, loss=0.0497]
Epoch: 18/25: 100%|██████████| 395/395 [00:17<00:00, 22.32batch/s, loss=0.0453]
Epoch: 19/25: 100%|██████████| 395/395 [00:23<00:00, 16.76batch/s, loss=0.045]
Epoch: 20/25: 100%|██████████| 395/395 [00:17<00:00, 23.00batch/s, loss=0.0424]
Epoch: 21/25: 100%|██████████| 395/395 [00:22<00:00, 17.66batch/s, loss=0.042]
Epoch: 22/25: 100%|██████████| 395/395 [00:19<00:00, 20.70batch/s, loss=0.0394]
Epoch: 23/25: 100%|██████████| 395/395 [00:18<00:00, 21.45batch/s, loss=0.0384]
Epoch: 24/25: 100%|██████████| 395/395 [00:19<00:00, 20.07batch/s, loss=0.0383]
Epoch: 25/25: 100%|██████████| 395/395 [00:18<00:00, 21.09batch/s, loss=0.0369]

----------------------------------------------
[System Complete: 0:10:50.239850]
```

Figure 10: Training procedure during 25 Epochs

# Appendix F Analysis and Discussion

| **Class** | **Accuracy** |
|---:|---:|
| 0 | 99.0% |
| 1 | 100.0 % |
| 2 | 64.8% |
| 3 | 97.6% |
| 4 | 100.0% |
| 5 | 94.2% |
| 6 | 96.5% |
| 7 | 78.7% |
| 8 | 100.0% |

Table 8: Accuracy per class