International Journal of Marketing and Technology
Vol. 12 Issue 02, February 2022,
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Predictive Data Analysisin Customer Profiling & K-Means Clustering

Pamela Nyatsine Email; Pam.farai@gmail.com

Abstract: Predictive analytics has grown over the years with many startups attributing their growth to this area of Big Data. The business performance KPIs can be scrutinized such as long term customer retention based off online analytics of the customers behavior with the company.
In recent months there has been concern over the use of machine learning to influence vulnerable groups and there is regulation in the pipeline in Europe to curb this. The future compromise would be a balanced approach to AI and machine learning using customer data and protecting the customer. However predictive analytics cannot be legislated away, while some have coined the term 'surveillance capitalism' to describe the manipulative nature in which Big Data analytics can be applied. Algorithms can allow companies to gain new insights into customer behavior.
Keywords: classification, targeted marketing, predictive analytics, k-means clustering

Introduction

'Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data'[1] . While knowing what transpired is good, the ability to predict what may happen in the future is even better. Data tools can assist us in moving from surplus data to information to knowledge and finally to action which can save the firm money, make it more profitable or impact the environment that the firm operates in for the good.
Predictive analytics can support entrepreneurial endeavors, however due to the nature of startups resources may not be available to engage in data analytics. Startups can save on resources by targeting the correct segment of society. They can gain new insights into customers' behavior through use of algorithms that monitor online activity e.g. mouse over. When using Power BI for example users can monitor, their startups via a mobile app showing a dashboard with all relevant statistics. The main justification would be competitiveness on a local, regional and in some cases international level.
Customers have evolved in tastes and preferences combined with use of social media and other technology they require more sophisticated marketing methods. In a physical store it is easier to know the customer unlike the digital world it is not always possible to know the customer. Users of a digital product may use pseudonyms as their username. They may or may not share location to identify where they reside. Algorithms can be utilized to segment the customers according to behavioral activity.

Background

Startups in Zimbabwe are competing on global, regional and local scale. To remain competitive, they require new skills and knowledge in order to acquire and retain customers in the long term. Big data analytics has made that possible with business intelligence profiling tools such as Tableau.There are some commonly used algorithms for classification such as k-means, Random Forest and Support Vector Machines. This study will review the k-means clustering algorithm.
Research within the context of machine learning covers a five stage process: firstly what needs to be done before the model is built, secondly how to reliably build the model, thirdly how to robustly evaluate the model, fourthly how to compare models fairly and lastly how to report results[2].
A model is a mathematical representation of a real world phenomena which can be employed to predict future events and outcomes [3]. It is possible to model a customer's online behavior such as whether they read product reviews, if they looked at product description before purchasing the product and many other online activities.

There are a number of models available for customer segmentation and profiling. One can combine a few of the models in order to achieve best results, this is known as ensemble methods. There are algorithms that can track mouse movements as one browses a website and therefore predict what items may be of interest to the customer as well as whether the customer will be a long term loyal customer. In education learning support systems have dashboards and can use machine learning to predict students' needs[4]. Startups in emerging markets owe a part of their success to predictive analytics. Big data has been a catalyst in startups growth [5]. There is development in Europe where there are proposals to limit use of AI in facial recognition, autonomous driving, online advertising driven by analytics from Tech giants based in US and China [6]. Some of the regulations would limit AI systems where they target specific vulnerable groups to influence their behavior.
One method is to perform customer segmentation using unsupervised learning (k-means clustering) to deliver better service to customers. In addition to the technical tools there are methods of profiling listed below[7]: the following profiling methods can assist in creating classes for modelling data before analyzing it. There are various classes that customers can be grouped in prior to predictive analysis being performed on the class.

Demographic Profiling
Business advisors advocate the use of a demographic profile to understand what kind of products appeal to a particular customer. Through demographic profiling, one can look into the customer's details such as geographical location, marital status, and educational qualifications to predict what product would appeal to the customer.

Cluster Coding
Class and social activities are closely linked, a middle income individual may likely live in a particular part of the city drive certain cars and engage in certain lifestyle activities. Through cluster coding the startup can target their marketing in a particular cluster.

Affinity Profiling
Affinity profiling allows one to study the buying habits of people in order to determine what kinds of products a particular customer needs. If a customer buys a camera on a website they may buy an SD card too. Using a tool such as Apriori one can suggest what products go together and market the related products.

Psychological Profiling
This tool can be used to tell a lot about a person by understanding what motivates them from a psychological perspective. If a customer wears brands mostly it would be safe to say they would be interested in high end brands that the company sells.

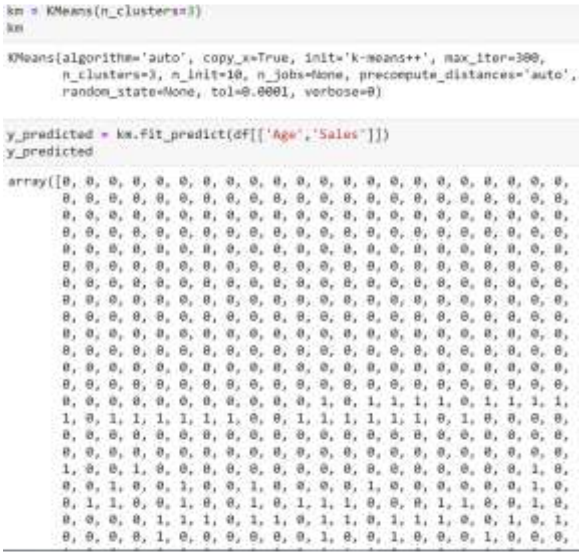
Lifestyle Coding
Through lifestyle coding a model can predict what a customer will be interested in based on their hobbies and habits. A collector of vintage cars may be interested in the accessories that go with those vintage cars.

Best practice is to experiment with multiple modelling tools and techniques which will highlight best practices for customer profiling. A solution should be customizable to suit the particular startups needs[3]. It is mandatory to utilize a number of methods of investigation in order to substantiate the results [8]

Definition of Key Variables

Indicator variables are included in the models as they represent real life features with a 0 or 1. This makes data easier to work with. The figure below shows an array in Jupyter Notebooks where the dataset has been converted to 1s and 0s representing the two clusters that the data was split into using k-means algorithm.

For certain analyses numeric values can be broken into classes. This a process of transforming data using indicator variables. The features in the model come from the columns in the dataset. String data can also be transformed into numeric values for ease of computation.



```
km = KMeans(n_clusters=3)
km

KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=None, tol=0.0001, verbose=0)

y_predicted = km.fit_predict(df[['Age','Sales']])
y_predicted

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0,
       0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0,
```

K-means clustering array 1

By employing predictive analytics it is possible to steer an organization in the desired direction and save on resources by not wasting in less beneficial endeavors[9]. Interactive digital dashboards allow exploration of events without technical knowledge as the software performs it all in the back end. Predictive analytics is a progressive and exciting area of big data analytics that is allowing productive use of the large volumes available due to the digitization of our modern lives. Harnessing the appropriate models will benefit both consumers and service providers. There is a learning curve with this new territory as there is no black and white.

Literature Review

The literature review highlights areas of ongoing research in which big data analytics with particular reference to predictive analytics. The majority of the literature is from published articles in 2021 with a few from previous years. Conference papers and books have also been consulted to give background to development of some of the current technologies that are in use. The focus is then narrowed to analysis of the proposed technology and specific area of study which is marketing. The emphasis on current literature as opposed to the past decade or so, is purely due to the fact that this area of research is a hot topic with new barriers being broken all around. The technologies are developing in real-time and many discoveries across various industries to underpin the importance of predictive analytics in aiding modern day businesses.The technologies are being employed around the globe from predictive analytics for a hospital in Zimbabwe, Human Resource predictive analytics for training requirements, to planetary geo science in Korea. Inductive reasoning is a data driven approach that draws general inferences from empirical data by observing patterns in the data [10]. A study in South Korea utilized Apriori algorithm to search for associative rules and keyword signals for wellness which is a current trend. The Apriori algorithm is exploratory in nature and useful in revealing relationships between entities, the association rules are based on a predetermined minimum confidence. There have been approaches to integrating business intelligence tools such as Power BI with machine learning using technologies such as Python [11]. Focus has shifted to use of algorithms in predictive models for future business planning.

Geosciences are utilizing big data analytics to observe patterns in planetary data for predictions on climate. The implementations are performed in the cloud using instances of virtual machines with pre-built data analytic tools [12].

The European Union is introducing some regulations on artificial intelligence which will affect tech giants in the US and China. This is being done in order to protect vulnerable groups in society

in a way that could affect or influence their behavior [6]. This includes surveillance advertising where an algorithm placed adverts on betting sites in front of gambling addicts.

The EU may require risk assessments of some of the applications with fines on a company's global sales as punitive damages for failure to comply. Users need to be notified of use of sentiment analysis to gauge their political orientation or users' emotions for example and many other features the application can collect from the user to use in marketing mix of the company's products and services. While this may discourage investment it protects users. There are digital rights analysts employed to work these features out. High risk AI systems may cause people to form an opinion or behave in a way they would not have done otherwise. AI systems can create biases or discrimination. Regulators are sometimes afraid to make many rules so they don't stifle innovation.

Big data analytics adoption in Nigeria has been slow according to a study by (Vincent 2021) and (U Prince)[13]. The study aimed to detect fraud using binary logistic regression. The study found that big data analytics protected against cyber-attacks. Big data aids in predicting future events such as threat detection of malware and phishing attempts. By analyzing large volumes of data and applying data models predictions are made accurately.

In Zimbabwe there hasn't been a much study of data analytics from a research perspective however there is much progress in the area of degrees being offered in data analytics and cloud computing this will increase the research base which has been small so far. A study done in 2019 by fellow students highlights how data analytics can be utilized in Zimbabwean hospitality industry [14].Chatbots can be utilized to collect customer data for further analysis after initial customer care [15].

Descriptive Analytics

Some of the existing tools in use today for Big Data Analytics include Hadoop, Cassandra and Spark[16]. Descriptive analytics examines data as it stands while predictive analytics first describes in order to understand and therefore predict a future outcome based on previous experiences. Data description gives understanding on the size and shape of the data. Descriptive analytics gives a narrative of the current and the past from a data set.

Predictive Analytics

Machine learning, deep learning and decision science are just some of the approaches for use in predictive analytics. Using algorithms a machine can learn and predict certain events to a certain level of accuracy. Association rule mining, logistic regression, time series analysis and Bayesian classification are among many of the machine learning methods available.

Predictive analytics can be employed influencing how teachers think about and teach students and in understanding education [17]. A study used AI-powered predictive analytics systems to tell a narrative about education through data dashboards. The data dashboards track social interactions among students, the resource utilization and assessments.

The use of predictive analytics has spread to Human Resources in employee turnover and training analysis requirements. Using software such as Power BI and machine learning models such as logistic regression Human Resource management can gain insight to benefit the business.

Prescriptive Analytics

While predictive analytics is useful in predicting a future event or outcome prescriptive analytics provides a specific course of action. It provides an answer to questions such as 'what should be done' using various technologies like neural networks, simulation and event processing [18]. Prescriptive analytics includes both predictive and descriptive analytics. The decision making is usually automated in models.

Detective Analytics

In manufacturing detective analytics can be employed to make diagnostics of collected data eliminate issues in the manufacturing that are picked through predictive analytics [19]. Smart manufacturing can be used to optimize customized products for customers.

Affective Analytics

Natural language processing has been instrumental in building the area of affective computing. Machine learning can check a user's emotional state [20]. The models rely on previous knowledge, lexical based approaches to provide accurate results.

Synthesis of Literature

Businesses have transformed from single transactions with clients to providing solutions in long term relationships through use of predictive and prescriptive analytics which is accomplished after descriptive analytics [21].Netflix worked with AWS to deploy storage in which they could store data that was detailed as much as possible about their customers, their algorithms work to recommend to customers content based on a host of factors. Netflix teamed up with Facebook to make their recommendation data more powerful [22]. Utilizing algorithms Netflix can influence what customers watch based on what they watched previously and interests from other platforms such as Facebook, using strong backend infrastructure.

The current literature highlights the strides that big data analytics is making with the use of specialized algorithms. There is room for growth in the use of predictive analytics as it is used across various industries. As discoveries are made results are published for all to benefit and build upon the progress which catapults the whole industry as knowledge is readily available. Researchers and practitioners alike will not need to re-invent the wheel as it were.

Business analytics is a process of collecting data, describing the data, analyzing and interpreting the data [23]. Analytic tools answer critical questions for business growth and continuity. Businesses have structured, semi-structured and hybrid data which can be mined for various outcomes.

Cloud-based Analytics as a Service (AaaS) is a growing trend [23]. A software vendor can provide access to an online platform for clients to perform data analytics for a monthly fee with an option to stop the subscription at any time. This option can be well suited for startups that are usually constrained on resources for setting up big data analytics technologies in-house. Technologies are changing rapidly hence investing heavily in current technologies may not be advantageous. By paying for the service and investing funds in other areas of the businesses that need the resources the most startups can have a leg to stand on to remain competitive in their various industries.

A hybrid approach of utilizing tools the business has on hand for data analytics while outsourcing the remainder of the analytics workload provides a favorable result for the startup. Tools that contain more costly licensing can be outsourced while more affordable or open source or in-house developed solutions can be run internally.

Predictive Analytics

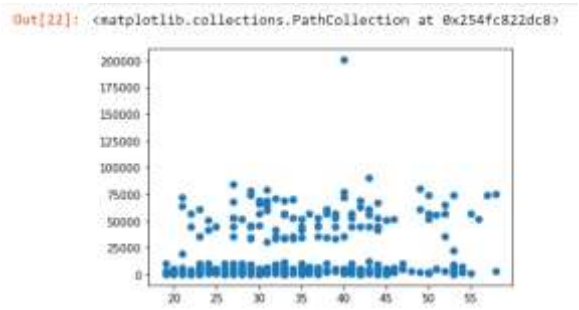
Recommender systems have a three pronged approach namely content based filtering, collaborative filtering or combination of both which is a hybrid [3]. Collaborative approach classifies customers in groups according to their tastes, preferences or other feature using an algorithm such as k-nearest neighbor. It uses two approaches that is item based and user based.

Data mining models can be used to recognize potentially profitable customers. A cross or up or deep selling classification model can show existing customers with higher purchasing power [24].

Data Understanding and Classification

In order to understand the data descriptive analytics are first applied to data. The size of adataset, the data types contained in the dataset, as well as the data shape are highlighted as a first step to data understanding before classification can begin. The type of data is grouped into attitudinal, behavioral, structured or unstructured, static or streamed. There is also demographic data. Before clustering data it can be put into a data matrix as a preprocessing step.

Data Classification is a method of data mining that is used in predictive models to uncover any hidden patterns in the datasets. As in the graph below the age of clients is plotted against their spending.



Scatterplot Sales and Age of Customers 1

It is used to predict any data category. A classifier is an algorithm that can classify potential clients into categories. When the unseen relationships are highlighted the model can predict outcome when new data is input. The classification algorithm predicts the class of the future event.

In marketing for instance a classifier can assist in targeted marketing campaigns. A data classifier can be built that can predict whether a customer will buy a product or not. Data classification can be a first step in building a predictive model when dealing with large datasets that can be tedious to work with at times. The data is partitioned into three categories according to the spending levels of the customer. Tier one two and three. Age of the clients assists with classification as well as the User ID/Email can assist with knowledge of Total number of clients. Assist with targeted marketing in increasing customer spend per head rather than looking at overall sales as an indicator of growth. Customer acquisition statistics can be tracked on a monthly basis. The delivery address can also be utilized as a unique identifier in some instances. Data clustering groups similar elements and data classification seeks to predict which class or category an element belongs to.Tools such as Google Colab, Gradio used for interfaces, Jupyter notebooks, datasets in csv format imported in Microsoft Excel for hypothesis testing template are useful in a classification ecosystem. There are a vast number of tools for serving a data model to predict trends in predictive analytics. The diagram highlights how a given predictive model works.

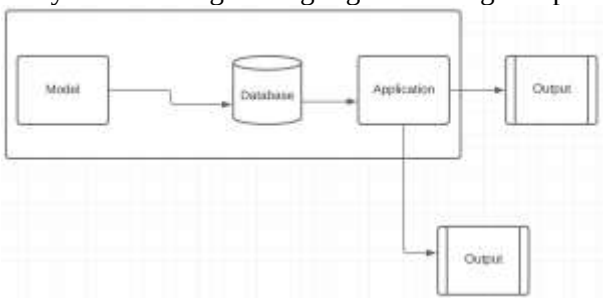
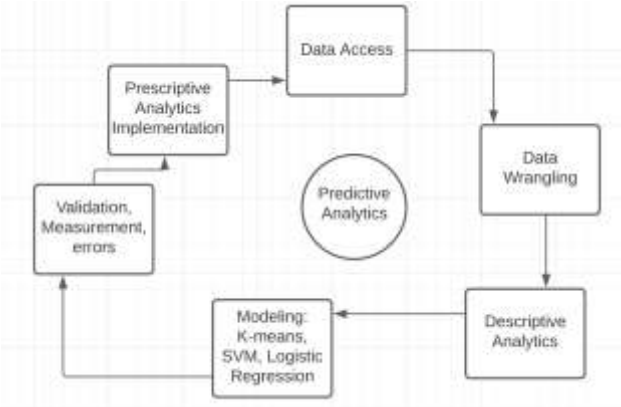


Figure 0-1 Method of Serving the Model

The data prediction is performed by building a classification based model. The steps involved are as follows: Collection of data about the past, current and potential customers from their interaction with the company's website. Next the transactional data from customers who actually purchase the products is collected. The training data is then selected which will be used to build the classification model. Some of the data is set aside for testing the model.

The model is tested until it is validated and the accuracy of its performance is verified on the historical data. When satisfied the model is deployed, as new incoming data for a given customer arrives the model classifies that customer as potentially fulfilling the requirement or not. Geocode to record the geographical coordinates of your criteria so you can predict by locations. Can use support vector machines, decision trees.



1.2 *Figure 0-2Predictive Analytics Flow Diagram*

Performance

Errors that may occur in research include, Population specification male and female consumers spend, binomial distribution using gender columnSampling and sample frame errorsInterpolate missing values age SelectionNon-responsiveMeasurement ROC measures the performance of the classifiers.

K-means Clustering for Customer classification

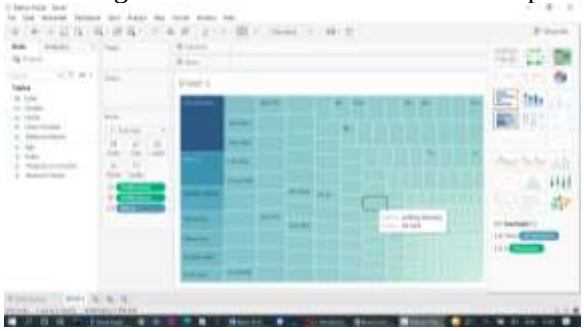
Classification aims to discover insights and patterns from data using different algorithms such as K-means algorithm. K is an input to the algorithm for predictive analysis; it standsfor the number of groupings that the algorithm must extract from a dataset, expressed algebraically as, k. A K-means algorithm divides a given dataset into k clusters and recalculates the new clusters' representatives. The user may predefine how many clusters the algorithm should put the data into e.g. k=3.The data in the figure below was divided into two clusters using the k-means clustering method.

```
df['cluster'] = y_predicted
df.head()
```

	Date	Name	Sales	Age	Gender	cluster
0	01/03/2021	Zee Mawere	3599.00	35	M	0
1	04/03/2021	Natalie Mill	2999.00	23	F	0
2	04/03/2021	Patrick Dodzo	3834.84	45	M	0
3	04/03/2021	Shadey White	497.96	29	F	0
4	06/03/2021	Bertha Anglo	5996.68	21	F	0

Figure 0-3 Dataset with clusters

The k-means algorithm is a useful for classification. When building new knowledge classification aids the process. In a supervised approach items can be assigned to certain classes while new interesting classes can be discovered in unsupervised learning [25].



In the above figure the largest client is in the wrong cluster as it is clustering according to total per transaction rather than total spend per client. In the figure below clients with total transaction of above RTGS20 000 are placed in cluster one while those with transactions below are in cluster 0. When comparing with the Tableau visualization one can see that the top client has been placed in cluster 0 as each transaction is below 20k while his total purchases are greater than even those of the clients placed in cluster 1. It is imperative to apply a number of tools and techniques before accepting any outcomes from a model.

```
df.tail()
```

	Date	Name	Sales	Age	Gender	cluster
682	31/08/2021	Dzi Dza	9288.06	32	M	0
683	31/08/2021	Leigh Benny	79180.67	31	F	1
684	31/08/2021	Never Lose	9000.84	32	M	0
685	31/08/2021	One Winner	60399.04	49	F	1
686	31/08/2021	Ta Zviona	90711.36	43	F	1

K-means algorithm is useful for clustering data. The k is an input to the algorithm which represents the number of groupings that the algorithm must extract from a given dataset. In the figure below k-means algorithm is applied to the age and sales data of a given startup.

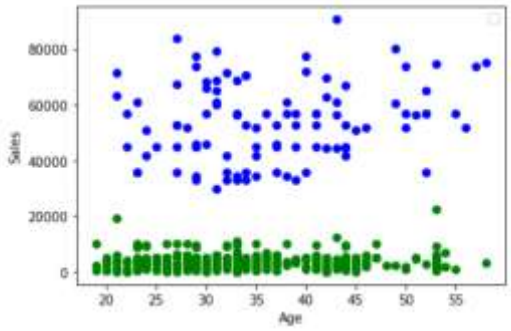
```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]

plt.scatter(df1.Age,df1['Sales'],color='green')
plt.scatter(df2.Age,df2['Sales'],color='blue')

plt.xlabel('Age')
plt.ylabel('Sales')
plt.legend()
```

```
No handles with labels found to put in legend.
```

```
<matplotlib.legend.Legend at 0x254fc850548>
```



The Euclidean distance is used to calculate the distance between two points. If the distance is small it means that the member belongs to that class.The clustering algorithm can be reiterated till the clusters do not change. A member of a cluster is the average or mean of all the items that are in the cluster.

Conclusion

Predictive models analyze data and predict future outcomes [3]. The ability to forecast the future outcomes in a given situation gives the distinction between business intelligence and predictive analytics.Classification-based predictor can be used to enhance targeted marketing and other practical applications of the business.

Classification is a sophisticated method of marketing to clients according to their classification bracket. K-means is a clustering method that provides an entry into the classification process. It has some limitations and so would best be implemented with another classification algorithm. Visualizations assist with presentation of data and show some information that may not be picked easily with application of an algorithm.

References

[1] SAS Institute Inc., "www.sas.com," SAS Institute Inc, 1 January 2021. [Online]. Available: https://www.sas.com/en_us/insights/analytics/predictive-analytics.html. [Accessed 27 June 2021].

[2] M. A. Lones, "How to avoid machine learning pitfalls: a guide for academic reserchers," *School of Mathematical and Computer Sciences, ,* pp. 1-17, 2021.

[3] A. Bari, M. Chaouchi and T. Jung, Predictive Analytics for Dummies, Hoboken: John Wiley & Sons, 2014.

[4] J. Jarke and F. Macgilchrist, "Dashboard Stories: How narratives told by predictive analytics, reconfigure roles risk and sociality in education," *Big Data and Society,* pp. 1-15, 2021.

[5] A. Behl, P. Dutta, S. Lessmann, Y. K. Dwivedi and S. Kar, "A Conceptual Framework for The Adoption of Big Data Analytics by E-Commerce Startups a Case based Approach," *Springer Link,* pp. 285-318, 2019.

[6] W. Knight, "Europe's Proposed Limits On AI Would have Global Consequences," *Fast Forward Newsletter,* pp. 1-6, 2021.

[7] Khera Communications Inc., "www.morebusiness.com," Khera Communiations Inc., 4 August 2021. [Online]. Available: https://www.morebusiness.com/understanding-customers/. [Accessed 4 August 2021].

[8] D. V. Thiel, Research Methods for Engineers, Cambridge: Cambridge University Press, 2014.

[9] D. White, "Business Predictive Analytics: Tools and Technologies," in *Data Analytics in Marketing, Entrepreneurship and Innovation*, London, Auerbach Publications, 2021, p. 192.

[10] Y. Lee, M.-L. Kim and S. Hong, "Big Data Analytics: Exploring the Well Being Trend in South Korea Through Inductive Reasoning," *KSII Transactions on Internet and Information Systems,* vol. 15, no. 6, pp. 1996-2011, 2021.

[11] V. Vashisht, N. Jahkmola, P. Manjarwar and N. Nikhil, "An Effective Approach for Integrating Microsoft Power BI Application with Python for Predictive Analytics," in *Micro Electronics and Telecommunication Engineering*, Singapore, Springer, 2021, pp. 469-477.

[12] A. M. Annex and K. W. Lewis, "ASAP-Stereo Automated Pipeline," in *5th Planetary Data and PSIDA*, Baltimore, 2021.

[13] T. Vincent and U. Prince, "Implementation of Critical Information Infrastructure Protection Techniques Against Cyber Attacks Using Big Data Analytics," *ResearchGate,* pp. 1-77, 2021.

[14] N. C. Shereni and M. Chambwe, "Hospitality Big Data Analytics in Developing Countries," *Taylor and Francis Online,* vol. 21, no. 3, pp. 361-369, 2019.

[15] M. P. Putra, "An Analysis of Big Data Analytics, IoT and Augmented Banking on Consumer Loan Banking Business in Germany," *Journal of Research on Business and Tourism,* vol. 1, no. 1, pp. 16-36, 2021.

[16] B. Jabir and N. Falih, "Big Data Analytics Opportunities and Challenges," *International Journal on Technical and Physical Problems of Engineering,* vol. 13, no. 47, pp. 20-26, 2021.

[17] J. Jarke and F. Macgilchrist, "Dashboard Stories: How Narratives told by predictive Analytics Reconfigure Roles, Risk and Sociality in Education," *Big Data & Society,* pp. 1-15, 2021.

[18] Gartner Inc., "https://www.gartner.com/en/information-technology/glossary/prescriptive-analytics," Gartner Inc., 1 January 2021. [Online]. Available: https://www.gartner.com/. [Accessed 23 August 2021].

[19] B. C. Menezes, J. D. Kelly, A. G. Leal and G. L. Le Roux, "Predictive, Prescriptive and Detective Analytics for Smart Manufacturing in the Information Age," *Science Direct,* vol. 52, no. 1, pp. 568-573, 2019.

[20] S. Gievska, K. Koroveshovski and T. Chavdarova, "A Hybrid Approach for Emotion Detection in Support of Affective Interaction," in *IEEE International Conference on Data Mining Workshop*, Washington, 2014.

[21] J. K. Von Bischhoffshausen, M. Paatsch, M. Reuter, G. Satzger and H. Fromm, "An Information System for Sales Team Assignments Utilising Predictive and Prescriptive Analytics," in *IEEE Xplore*, Lisbon, 2015.

[22] E. Zabalawi and A. A. Jammal, "Innovation Analytics," in *Data Analytics in Marketing, Entrepreneurship and Innovation*, New York, CRC Press Taylor & Francis Group, 2021, pp. 15-29.

[23] G. C. Deka, "Big Data Predictive and Prescriptive Analytics," in *Handbook of Research on Cloud Infrastructures for Big Data Analytics*, India, IGI Global, 2014, pp. 1-22.

[24] A. Chorianopoulos, Effective CRM using Predictive Analysis, West Sussex: Wiley Publishers, 2015.

[25] J. P. Mueller and L. Massaron, "Clustering," in *Python for Data Science for Dummies*, Hoboken, John Wiley & Sons Inc, 2015, pp. 273-288.

[26] Postal and Telecommunication Regulatory Authority of Zimbabwe, "http://www.potraz.gov.zw," December 2015. [Online]. Available: http://www.potraz.gov.zw. [Accessed 17 August 2021].

[27] T. A. Hamid, M. Adnan, A. Qaiser, M. Rehmani, A. Bhattarai and R. T. Naveed, "Advances in Supply Chain Management using Big Data Business Analytics," *International Journal of Innovation, Creativity and Change,* vol. 15, no. 8, pp. 807-819, 2021.

[28] E. Mnif, I. Lacombe and A. Jarboui, "Users' Perception Towards Bitcoin Green with Big Data

Analytics," *Emerald Insight* , pp. 1-25, 2021.

[29] M. Machikiche, "Research Methods Presentation," 2019.

[30] A. Bari, M. Chaouchi and T. Jung, Predictive Analytics for Dummies, New Jersey: John Wiley & Sons Inc, 2014.