Movie recommendation system

Name of student

Name of professor

University

Course

Date

**Abstract**

Data has been the core center of computer operations and applications for a while now. With the inception of web 2.0. Data became the core interest of most of the applications that were in use. The use of CRUD operations and the need to find meaningful responses from data has been of Companies want to know how well they are doing, but this level of analytics and knowledge can only be established based on the real insights that data gives.

Further, the evaluation of data systems has seen the adoption of big data in many of the application systems at industrial level. Big data is data that is fast moving, comes in many forms and is voluminous. With the increased usage of social media interactions, online presence and ubiquitous computing, many developers and data scientists are now shifting to find the need in intelligent data.

Intelligent data is data that has learnt from its past. It uses its past to gauge its future. This level of analytics is very key to the sales and marketing teams so as to establish which targeted campaigns to offer and in which specific regions this level of campaign is most likely going to be effective.

Consequently, we can say that big data, machine learning, robotic automations are the gold of the future in the tech industry. In this particular study, the researcher explored how this is similar in a typical movie recommendation system. Based on data gathered from the end users, data was cleaned, analyzed, and algorithms applied to it so that insightful conclusions cold be arrived at.

**Introduction**

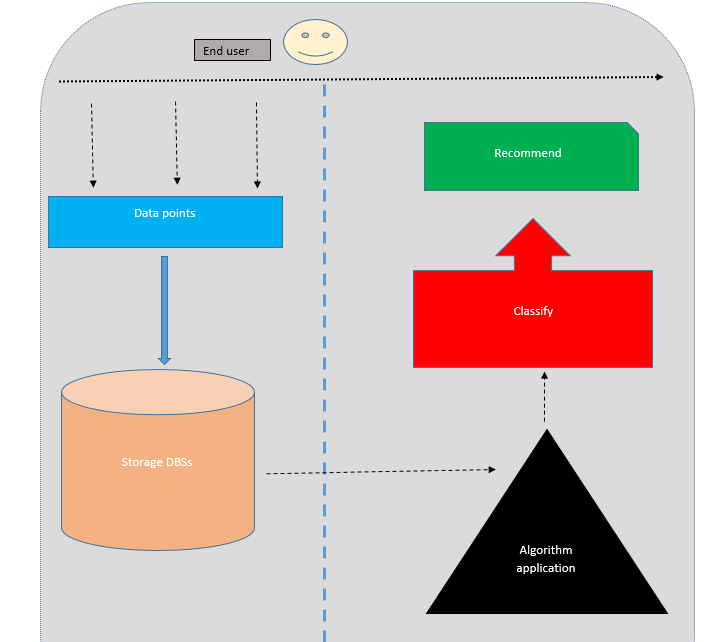
Before predictive analytics, a lot happens to the data and information presented to the end user. The initial process usually include data collection. During this process, the data is gathered from the data interaction points that might include ERP systems, social media posts, online search experiences, call logs or even form submissions. The next step involves the storage of these data. Usually structured database systems are very good tat this, since they give the ability to analyze and present data.

Structured database applications such as Oracle and IBMSs DB2 are very effective in handling structured data. Another concept that is quietly implemented in the background during this process is the business intelligence application also called the BI. The BI already has some knowledge based on its existing algorithms and code packages. These are utilized in the process of trying to affirm the measures of central tendencies and the trends associated with such data.

Moreover, other developers have gone ahead to implement algorithms that can be utilized in large scale sales and marketing organizations. The idea here is to be able to recommend to the user the final stake of product and other related product based on their experiences. This is the reason why online users are able to see other related searches based on the search experiences already gathered and implemented. This may happen based on complimentary products of other products.

Let’s pick a quick similar case of Netflix. A user searches for a movie say “Squid Game”, which is a Korean movie, once the user finishes watching this movie, a similar Korean movies might be on their to watch next list, since the movie application algorithms is trying to implement the idea that if the user loves Korean Movies, so they will always continue watching other related Korean movies. This is the way product recommendation algorithms work.

The process architecture is as follows:



From the above architecture it can be seen that process of recommending an item to the user involves 5 process which must all be passed. This is where machine learning comes in. This is what the researcher used in coming up with a movie recommendation system. Let’s observe the case below on how this is achieved.

**Movie Recommendation system.**

In a typical movie recommendation system, we assume that the movies are listed in some website or web application like say Netflix or Amazon Prime, each movie belongs to a particular genre like say comedy, thriller, war or adventure horror. Subsequently, when a user comes to watch a movie and say selects British comedy, the user process to watch the movie till end of the move, and once they are done, the algorithm behind the scenes, checks the initial movie that the user watched and tries to get its classification and based on that genre type classification, the algorithm is able to recommend other movies in the same genre since its understanding is that that is what the user loves and should therefore be watched by the user. In most cases these recommendations turn out to be above 80% accurate.

**False positives;**

A false positive is a scenario in which a condition seems to exist, whereas it does not. During recommendations, this is the margin of error that is usually left out for data scientist to evaluate and consider. Not all recommendations are based on a true input. Some of the inputs could be by accident, or wrong location, or merely unintended searches.

Recommendation runs on the backbone of prediction. The final goal of all model is to be able to predict the outcome based on the evaluated inputs of the given data and scenario. This is what the researcher establishes in this report.

**Problem statement**

Many organizations are usually faced with multiple problems when it comes to making sense from their data. Traditional analytics has not been quite effective though. Aside from just the summary of the existing counts and figures, one of the greater changes lies in the ability of the existing analytical tools to predict to the existing customer and be able to know the behavior or the consumer patterns of such users. With this in mind, companies are able to establish the required cause of actions for such customers so that they are able to improve their search experiences online.

In a similar context, the researcher sought out to establish a movie recommendation system for customers based on the supplied data of its users. The issue at hand is the ability to predict the move that the user should be watching next based on their current inputs and selections.

**Aims and objectives**

The aims and objectives of this project are as follows:

1. To identify the trend within the movies dataset
2. To classify the related dataset variables
3. To develop a recommendation based on the supplied dataset

**Literature reviews**

Structured database system are the most commonly used databases. They are used to store data across commercial and small scale business applications Elmasri, & Navathe (2011). They support common query languages for retrieving and manipulating data across database applications.

There usage has however grown just beyond the common tabular form of presenting data to the now unstructured database applications. Unstructured databases are the new norm in handling data types of several inputs, For instance, there is no specified column for a given data type or input, each and very data tope can fit in any given column.

According to Mueller & Massaron (2021), there is a very close relationship between machine learning and artificial intelligent systems. Most artificial intelligent systems actually implement machine learning algorithms behind the scenes. There are a number of conditional statements and rules that have been written on the backside of the robotics. Usually a program algorithm/code is written and inserted into the memory of the AI agent, then this code is programmed onto the machine of this application, the robotic machine will then read from its own memory and implement the algorithm that the owner wants it do.

However, this study does not establish the risks associated with AI and its implications on the work force since a lot of people are afraid that the automation of some of these agents could lead to job replacements

Machine learning can also be used in predictive analytics as indicated by Petropoulos et al (2020). In a study done among some of the top 50 banks in the United States, it was easier to classify using Random forest techniques the possibly of banks being in-solved. The data derived from this study took a sample of some of the banks across the US that had had, a history of insolvency, like their locations, customer categories, book of accounts, net profit margins and other government regulatory factors such as taxation and the legal frameworks surrounding the establishments and operation of such banks.

The study established that it was much easier to relate the factors to the possibility of banks being acquired or not. This was done based on the measure of the correlation of the factors in play. Where strong positive correlation exists between these factors, then the prediction was certain and was due to happen. This study however does not discuss some of the external factors or reasons such as customer behaviors, revenue earnings of the customers and the prevailing economic conditions of the areas.

In another study done by Fanca et al (2020), there is an illustration between the user input and the information filtering methods used. The study goes ahead to give movies, restaurants, hotels and reservation systems an option to select from the various options supplied by the recommending system. The filtering method is a subset of the larger supporting algorithm that narrows down to the selected sub-class of the algorithm. This study concludes that, this is typically how most of this applications work by fileting this information for the end users by leaving out the larger chunk of billions of metadata that is out there in the search.

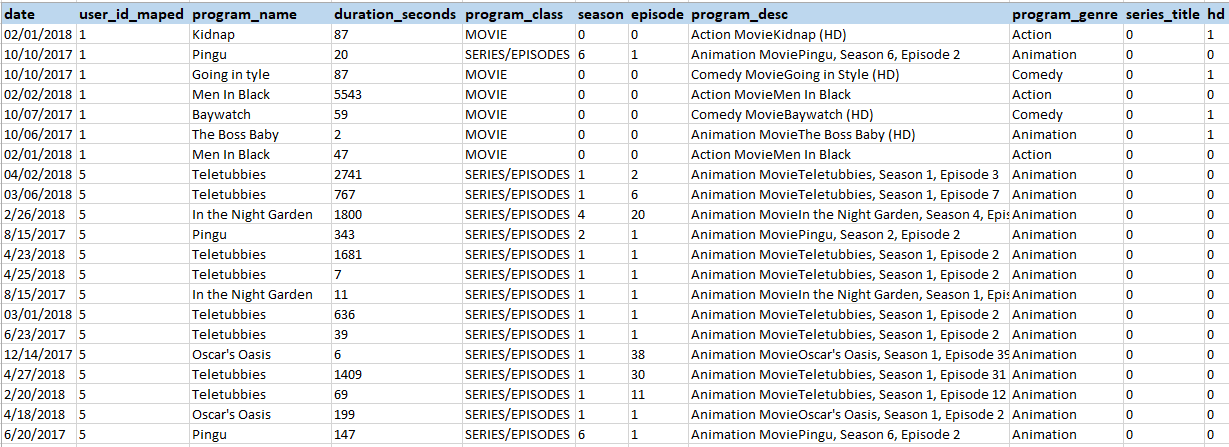
Further, Furtado (2020), illustrates the difference in content based recommendation visa a Vis the system based recommendation. Content based recommendation is still so primary and basic, not so widely used and embraced. This approach is manual and tiring, based on eye balling and this is the reason why these study establishes the need for automated system approaches and its applications on movie recommendations.

**Research Methodology**

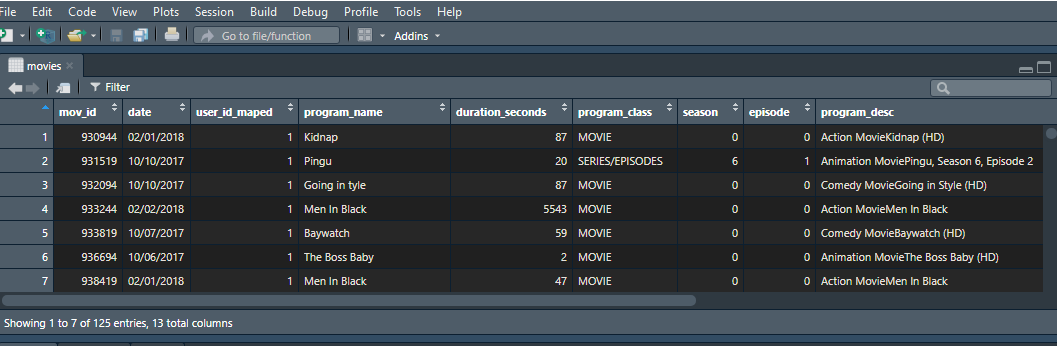
The researcher gathered data from several movie data points. The language for analysis in this project is R programming. The IDE used for this analysis is R-studio development environment.

**Dataset description**

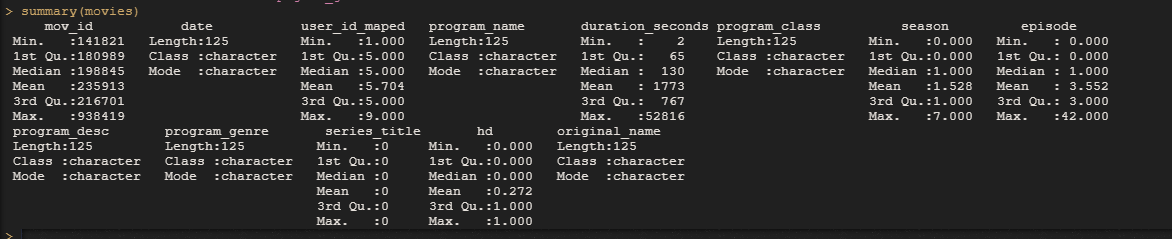
The dataset selected for this study contains a total of 126 records from which the researcher applied machine learning algorithm in order to try and establish the classification and recommend a movie approximately. A quick preview of the dataset looks as below:



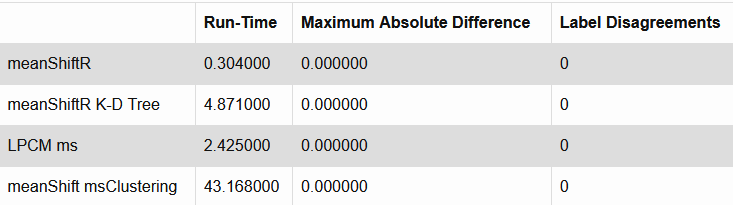
The next step involved loading the dataset into R-studio developer:

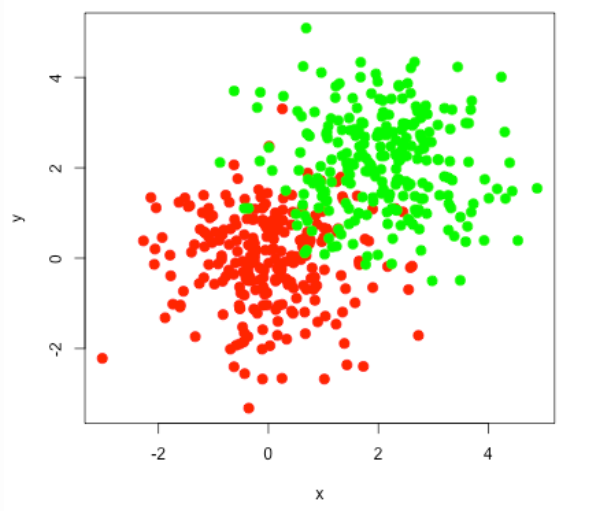


The next step is to get a statistical summary from this dataset;



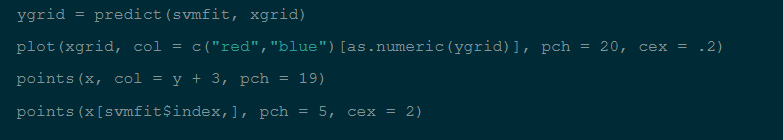
**The MeanshiftR Algorithm**



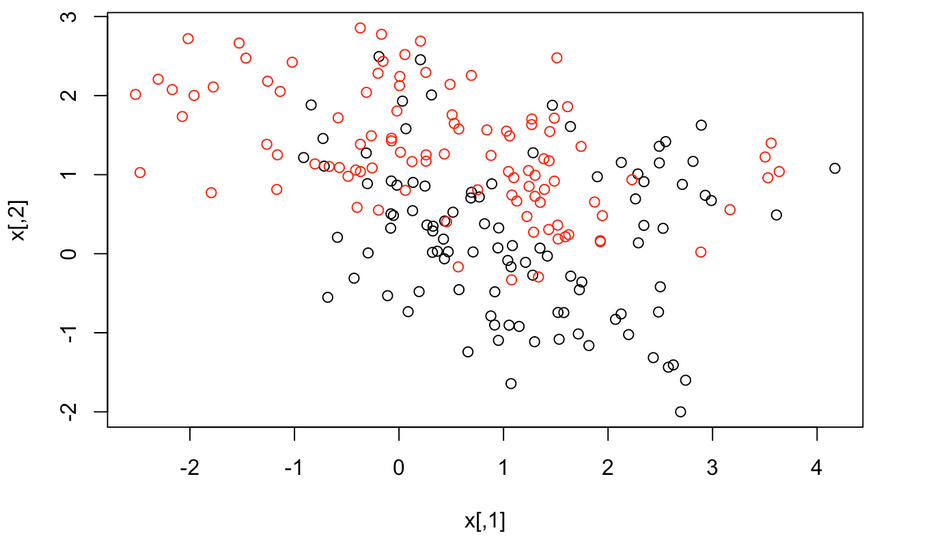


As seen from the visualizations of the model, the relation is not quite string in among the datasets, this gives the researcher an almost accuracy of 40%, which means that in trying to establish a classification of the movie object, chances are 0.4.

**The Support Vector model algorithm**



Visualize the plot



Again, the relationship between the movie genre type and the movie name distribution as shown above. The classification of a single movie based on the provided genres is likely to yield an accuracy of say 0.39%.

**Summary and conclusions**

As seen and observed, recommendation systems are as effective as the model and algorithm that handles them. Also, based on the analysis done, the accuracy of the MeanshiftR algorithm seemed higher in due its correlation unlike the supervised vector model. The higher the positive correlation the higher the accuracy of the model.

**References**

Elmasri, R., & Navathe, S. B. (2011). *Database systems* (Vol. 9). Boston, MA: Pearson Education.

Mueller, J. P., & Massaron, L. (2021). *Machine learning for dummies*. John Wiley & Sons.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, *36*(3), 1092-1113.

Fanca, A., Puscasiu, A., Gota, D. I., & Valean, H. (2020, October). Recommendation Systems with Machine Learning. In *2020 21th International Carpathian Control Conference (ICCC)* (pp. 1-6). IEEE.

Furtado, F., & Singh, A. (2020). Movie recommendation system using machine learning. International journal of research in industrial engineering, 9(1), 84-98