International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 09 | Sep 2020 www.irjet.net p-ISSN: 2395-0072

Novel Genre Classification using Deep Learning

Nishkala G[1], Bhavya P R[2]

[1]*Student, Department of BCA, M.O.P Vaishnav College for Women, Chennai, TamilNadu, India.*
[2]*Student, Department of BCA, M.O.P Vaishnav College for Women, Chennai, TamilNadu, India.*

Abstract - *The goal of this paper is novel genre classification by using the concepts of deep learning. Deep learning has machine learning algorithms which uses multiple layers to extract various higher qualities and features from raw data. There are various machine learning algorithms such as kNN(k-Nearest Neighbour), LVQ(Learning Vector Quantization), SOM(Self-Organising Map), LWL(Locally Weighed Learning). This paper mainly discusses about the classification of the novels in terms of their genre such as comics,educational,spiritual,crime etc. and the goal is to predict which genre is preferable to the readers and the public.*

Key Words: **kmeans, Matplotlib, Genre classification, Deep Learning, Algorithm.**

1. INTRODUCTION

Novel genre classification refers to classifying novels based on their genres like romcom, comedy, crime and mystery and analysing which is the most preferred genre by the public using the concept of Deep Learning. Deep learning which is also called as hierarchical learning is a branch of machine learning which uses layers to extract high level information from raw data and presents hierarchical representations of data. Most modern deep learning models are based on artificial neural networks. In Deep Learning each level transforms the obtained data into a higher level. This can be supervised, semi supervised and unsupervised. Deep learning algorithms mainly applicable to unsupervised learning tasks. The goal of these neural networks was to solve problem in the same way as a human brain does. Artificial Neural Network consists artificial neurons which are similar to the neurons in the human brain. Deep Neural Network is a type of ANN with multiple layers. The first layer processes the raw input data to a higher level. The second layer processes the data to a even more higher level by including the IP addresses. The third layer processes data obtained from second layer and adds more information like the geographic locations etc. and this process continues until the data is completely processesd. Deep Learning is a benefit as the volume of unlabelled data is higher than that of labelled data. Hence it provides a easier method of analysing and processing larger volumes of unstructured data.Deep learning is used in areas such as recognising images, recognising audios and videos.

In this research paper the concept of deep learning is used to analyse the data collected from public and predict the most preferred genre of novels. The software used is TensorFlow. From this type of analysis it becomes easier for various amateur authors as they are able to know the public preferences.

1.1 Literature Review:

From a survey conducted in 1973 by Norvell, it was reported that age and gender are two important factors in terms of reading habits. From a survey conducted by McKenna, Kearn, & Ellsworth, 1995; Clark & Foster, 2005; Shafi & Loan, 2010 it was reported that females enjoy reading more than males. From a survey conducted by Clark (2012) it was found out that 56.7% of girls enjoy reading 43.8% of boys enjoy reading. Devarajan (1989) and Tella and Akande (2007) stated that reading novels is mostly the student's first choice. Wicks (1995) confirmed that boys aged between 13 and 15 prefer reading fiction novels over non-fiction.Vakkari and Serola (2012) took a survey of 1000 adolescents and adults and found out that 70% prefer fictional books,67% prefer non-fictional books. Walia and Sinha (2014) reported that fiction is most preferred and the various genres of fiction that were most preferred is thrillers (35.9%),horror (22.4%) and romance 18.4%. From the survey it was also found out that girls preferred novels and serial books whereas boys preferred thrillers and actions. 70% of females preferred romance, 52% of femles preferred horror. 56% of males preferred war-spy stories, 45% preferred crime and mystery. According to a study research author use various tactics to attract readers and analyzing the most preferred genre by the public will be helpful for the authors. The software used to analyse the data is Anaconda.

2. Analysis of Data

Matplotlib:

```
Out[39]: [<matplotlib.lines.Line2D at 0x187c1d6fa88>]
```

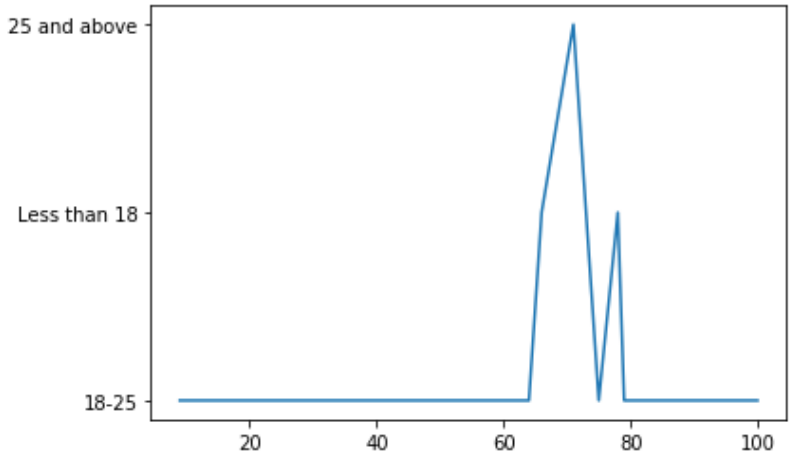

Figure 2.1: Analysis of Romcom Genre with age as constraint.

In the above image, the analysed genre is romcom. The genre is analysed with the age group of people reading that particular genre. The x axis represents the range. The y-axis represents the readers age. From the above graph, it is analysed that 70 percent of the readers of the age group 25 and above read romcom novels.80% percent of the readers are between the age group of 18 and 25. The maximum readers of romcom novels are between the age of 25 and above.

```
Out[67]: [<matplotlib.lines.Line2D at 0x187c31a4e88>]
```

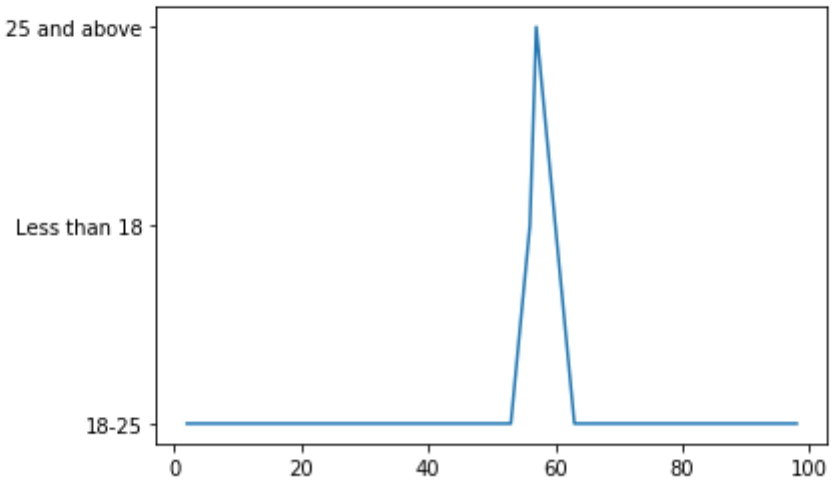

Figure 2.2: Analysis of Crime Genre with age as constraint.

In the above image, the analysed genre is crime. The y-axis represents the age group of people reading crime novels. From the above graph we get to know that about 60 percent of people reading crime novels are of the age group which is 25 and above. Hence the maximum readers of crime novels are within the age group of 25 and above.

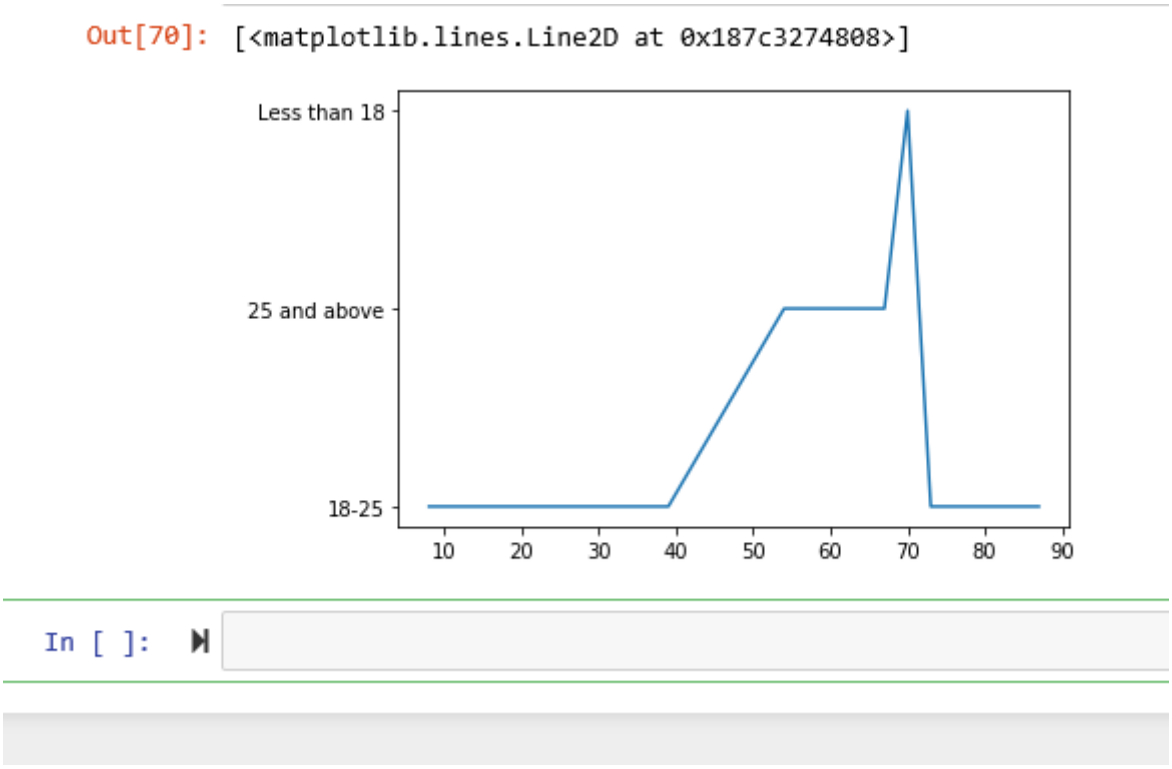

Figure 2.3: Analysis of Comedy Genre with respect to age.

In the above image the genre analysed is comedy. The y-axis represents age group. From the above graph we analyse that 75 percent of comedy novel readers are of the age group less than 18 years.55 percent of comedy novels readers are of the age group 25 and above. Hence the maximum readers of crime novels are within the age group of less than 18 years.

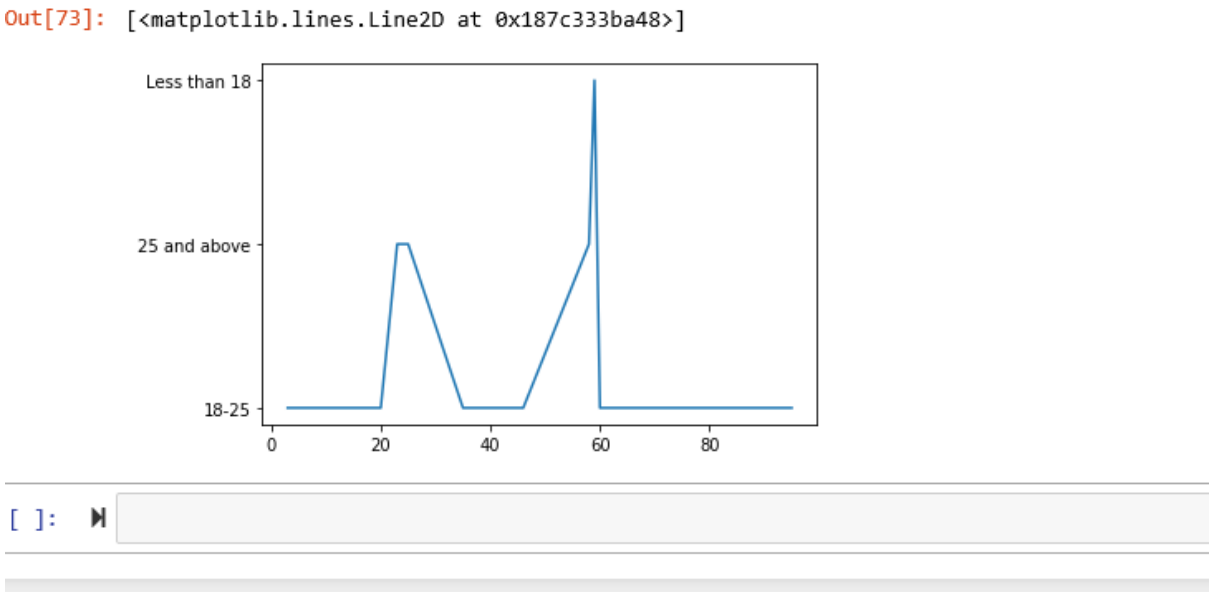

Figure 2.4: Analysis Of Horror Genre with age as constraint.

In the above image the analysed genre is horror. The y-axis represents the age. From the above graph, it is analysed that 60 percent of the readers are of the age group less than 18 years. 30 percent of the readers are of the age group 25 and above years. Hence the maximum readers of the horror novels are within the age group of less than 18 years.

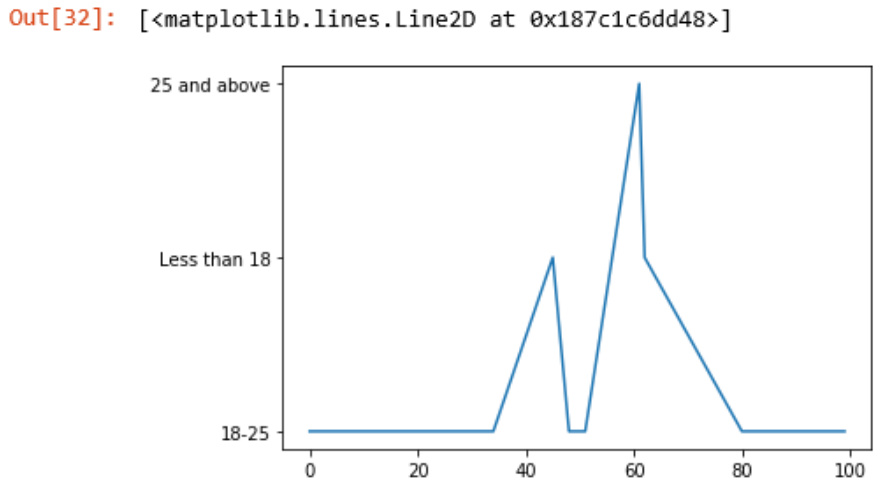

Figure 2.5: Analysis of Psychological Thriller Genre with age as constraint.

In the above image, the analysed genre is psychological thriller. The y-axis represents the age group. From the graph, it is analysed that 60 percent of readers of the age group 25 and above prefer psychological thriller novels. 40 percent of readers are of the age group less than 18 years. Hence the maximum readers of psychological thrillers are of the age group 25 and above.

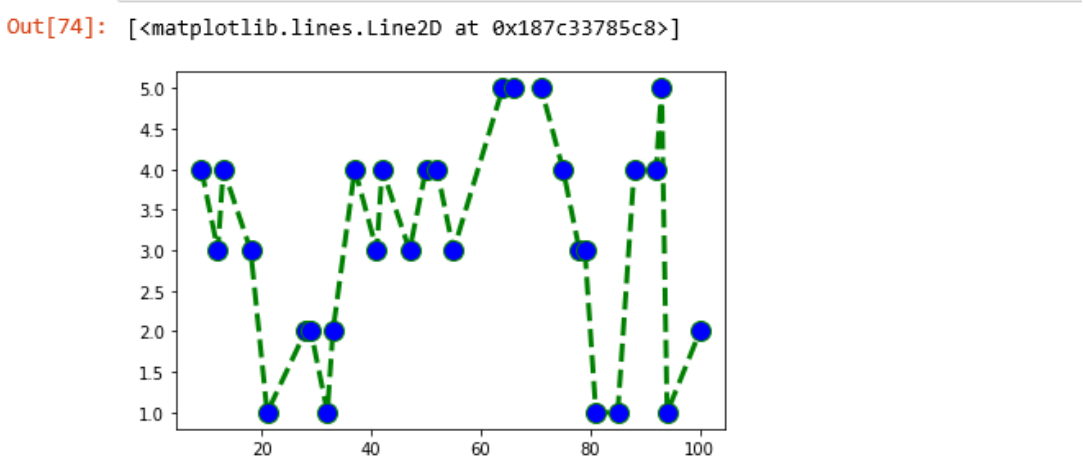

Figure 2.6: Analysis of Romcom Genre with range of enjoyment readers gain as constraint.

In the above image, the analysed genre is romcom and the constraint that is being analysed is the range of enjoyment reader gain by reading this genre of novels. From the above graph the Y-axis represents the range of values from 1 to 5 where 1 represents the readers do not enjoy while reading this particular genre and 5 represents readers enjoy reading this novel very much. From the above graph, it is analysed that 90 percent of readers enjoy reading the romcom novel.

```
Out[75]: [<matplotlib.lines.Line2D at 0x187c2e1aa88>]
```

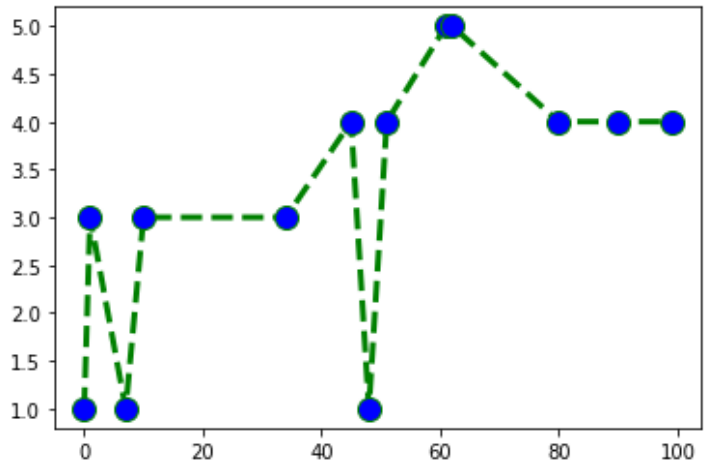

Figure 2.7: Analysis of Psychological Thriller Genre with range of enjoyment readers gain as constraint.

In the above image, the analysed genre is psychological thriller. The y-axis represents the range of value from 1 to5 where 1 represents the readers do not enjoy reading this genre of novel and 5 represents readers enjoy this genre very much. From the above image it is analysed that 70 percent of readers enjoy reading this genre, 50 percent of readers do not enjoy reading this genre and 35 percent of readers are average enjoyers of this genre. Hence the maximum readers enjoy reading this novel.

```
Out[76]: [<matplotlib.lines.Line2D at 0x187c30fff48>]
```

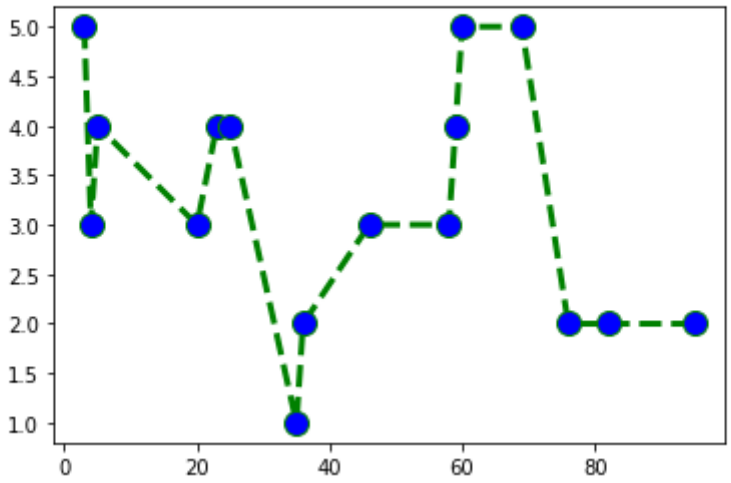

Figure 2.8: Analysis of Horror Genre with range of enjoyment readers gain as constraint.

In the image above, the genre being analysed is horror. From the above image, it is analysed that 70 percent of readers enjoy reading this genre. 60 percent enjoy very little by reading this genre.35 percent of readers do not enjoy reading this genre. 21 percent of readers enjoy reading this genre but are not completely enjoying it.80 percent of readers do not enjoy reading this genre in some parts, but enjoy reading it due to some parts. Hence it is analysed that most readers enjoy very little by reading this genre.

```
Out[77]: [<matplotlib.lines.Line2D at 0x187c310cb48>]
```

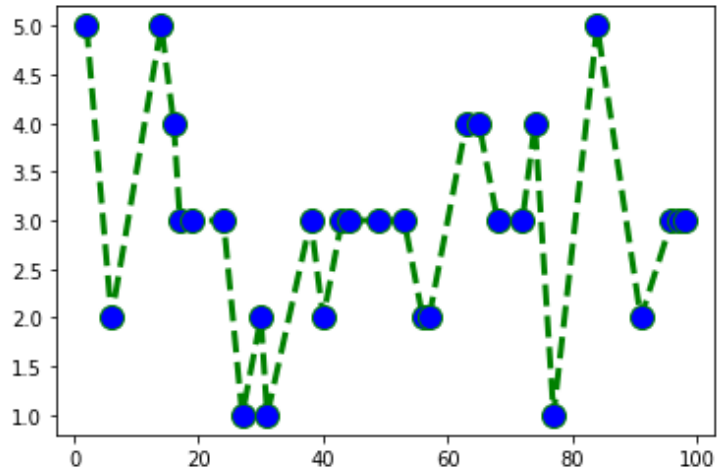

Figure 2.9: Analysis of Crime and Mystery Genre with range of enjoyment readers gain as constraint.

In the above image, the genre being analysed is crime and mystery. From the above image 10 percent of readers enjoy some parts of this genre and do not enjoy most of the parts, 50 percent of readers enjoy some parts and do not enjoy some parts of this novel. 79 percent of readers do not enjoy reading this genre. 85 percent enjoy reading this genre completely. 90 percent of readers enjoy some parts and do not enjoy some parts of this genre. Hence it is analysed that most the readers have given a 3 on scale of 5 which means that they enjoy reading this genre but do not enjoy completely.

```
Out[78]: [<matplotlib.lines.Line2D at 0x187c340ad08>]
```

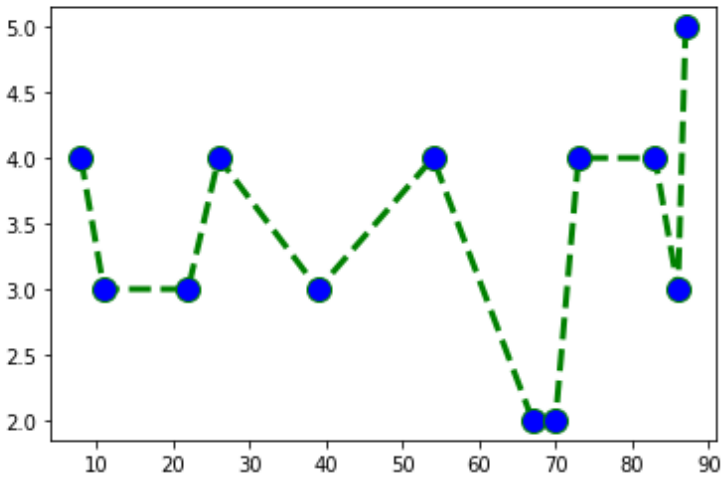

Figure 2.9: Analysis of Comedy Genre with range of enjoyment readers gain as constraint.

In the above image, the genre being analysed is comedy. From the image, it is analyzed that 40 percent enjoy some parts and do not enjoy some parts. 60 percent enjoy most of the parts but do not enjoy little parts of this genre. 70 percent do not enjoy reading this genre. 85 percent enjoy most of this genre. 89 enjoy reading this genre completely. Hence most of the readers enjoy reading this novel.

Using k-means:

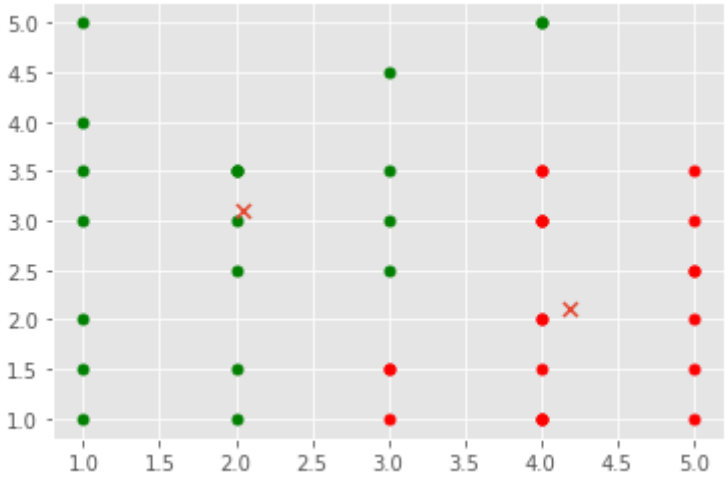

Figure 2.11: Novel genres classified using k-means.

In the above image, the novel genre is classified using k means algorithm. The x-axis represents the various novel genres, 1-Psychological Thriller, 2-Crime and mystery, 3-Comedy, 4-Horror, 5-Romcom. These are various genres chosen for classification. The y-axis represents the rating on a scale of 1 to 5.k-means divides the given dataset into clusters and mines them. The green cluster represents novels of 1 ,2,3 genres and red cluster represents for 3,4,5 genres. From the above figure, it is analysed that for the genre psychological thriller, the maximum rating on a scale of 5 is 5/5, for crime and mystery it is 3.5/5,for comedy it is 4/5,for horror it is 5/5 and for romcom it is 3.5/5. Hence from kmeans it is concluded that readers enjoy reading psychological and horror novels the most.

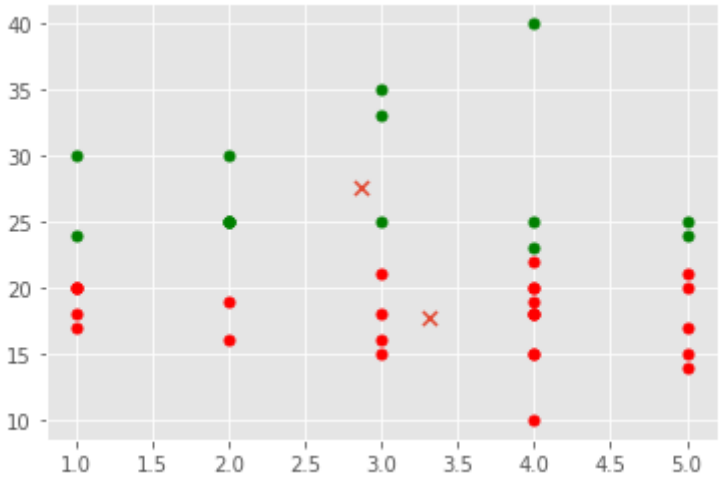

Figure 2.12: Novel genres classified using kmeans with respect to age of readers.

In the above image, the novel genre is classified using k means algorithm. The x-axis represents the various novel genres,1-Psychological Thriller, 2-Crime and mystery,3-Comedy,4-Horror,5-Romcom. These are various genres chosen for classification. The y-axis represents the age of various readers. k-means divides the given dataset into clusters and mines them. The green cluster represents the age which is above 25, and represents the age below 25. From the above figure ,it is analysed that for the genre psychological thriller, is read by people between age of 17 and 30. The genre crime and mystery is read by people between the age of 16 and 30. The genre comedy is read by people between the age 15 and 35. The genre horror is read by people between the age of 10 and 40. The genre romcom is read by between the age of 14 and 25. Hence using kmeans, the preference of the genre to different age groups of people are analysed.

3. CONCLUSIONS

From the above analysis,the novels are classified according to various genres and compared with constraints like the age and the enjoyment of reading that particular genre. Each genre is analysed using the two constraints and analysis is done to compare between the genres and find out the genre which readers enjoy.The analysis is also performed using kmeans algorithm and the genre are classified according to the age groups. The analysis for the enjoyment of reading the genre is performed using kmeans and plotted using matplotlib. According to kmeans algorithm, readers enjoy reading horror and psychological thriller novels the most.

From the above analysis, the different genres are classified against various constraints like the enjoyment of reading this genre on a scale of 5, the age constraint by which it is analysed which age group prefer reading which kind of genre. The various methods used is matplotlib for plotting graphs and kmeans algorithms. Each genre is analysed and compared with each other to find out which genre is most preferred to the readers. Using this analysis data, amateur authors have a idea to select the genre which is most preferred by the readers and write their novel in that novel to make it a big success. Authors also get an idea of age group of readers preferring a particular genre.

3.1 Inference:

From the analysis it is inferred that on a scale of 5, readers enjoy reading horror and psychological thriller the most and enjoy reading the crime and mystery, romcom genres the least. The comedy genre is enjoyed by most of the people and on a scale of the rating is 4/5. It is also inferred that different age of group of people like reading different genres.

REFERENCES

[1] https://www.youtube.com/watch?v=ZNWQN_g_ZsI&feature=youtu.be

[2] https://www.geeksforgeeks.org/graph-plotting-in-python-set-1/

[3] http://benalexkeen.com/k-means-clustering-in-python/

[4] https://www.geeksforgeeks.org/analysis-of-test-data-using-k-means-clustering-in-python/

[5] https://www.datacamp.com/community/tutorials/k-means-clustering-python

[6] https://realpython.com/python-matplotlib-guide/