3/3/2025

(Student C00270632) Qadeer Hussain

Lecture: Greg doyle

Data Science & ML Log

# **Bayesian Classifiers Log**

**17/01/2025**

* Reading up on Bayesian Classifiers notebook.
* Researching dataset to use.
* Dataset link: <https://archive.ics.uci.edu/dataset/2/adult>
* Predict whether annual income of an individual exceeds $50K/yr based on census data.

**23/01/2025**

* Reading up more on Bayesian Classifiers.

**18/02/2025**

* Decided on doing Gaussian Bayes Classifier as data is continuous numerical data.
* Worked on implementing Gaussian Bayes.
* The dataset contained both continuous and categorical variables which is why Gaussian was chosen.
* However Multinomial Bayes Classifier was also implemented to test to improve the performance as you can see in the images below.

**Gaussian Bayes Classification Report**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Multinomial Bayes Classification report**

**A screenshot of a computer

AI-generated content may be incorrect.**

**26/02/2025**

* Added confusion matrix on to the Multinomial Bayes Classifier.

**Multinomial Bayes Confusion Matrix**

A diagram of a confusion matrix

AI-generated content may be incorrect.

**03/03/2025**

* Made read me which includes the following:
  + Project Description
  + Data Source
  + Data
  + Processing
  + Data Understanding and Visualisation
  + Algorithm
  + Online Sources
  + Tools and Tech Used

# **Support Vector Machine Learning Log**

**26/02/2025**

* Read up on Support vector machine notebook
* Research dataset to use.
* Dataset Link: <https://archive.ics.uci.edu/dataset/80/optical+recognition+of+handwritten+digits>
* Predict handwritten digits from images

**27/02/2025**

* Reading up on more SVM.
* Attempt to implement SVM Model.

**SVM Classification Report**

A screenshot of a computer

AI-generated content may be incorrect.

**SVM Confusion matrix**

A graph of blue squares

AI-generated content may be incorrect.

**Model in use**

A collage of images of the letter h

AI-generated content may be incorrect.

**03/03/2025**

* Made read me which includes the following:
  + Project Description
  + Data Source
  + Data
  + Processing
  + Data Understanding and Visualisation
  + Algorithm
  + Online Sources
  + Tools and Tech Used

# **K Nearest Neighbour**

**27/02/2025**

* Read up on K Nearest Neighbours
* Research dataset to use
* Dataset Link: <https://archive.ics.uci.edu/dataset/19/car+evaluation>
* Predict the classification of a car based on multiple factors
* Implement KNN model

**KNN Classification Report**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**KNN Confusion Matrix**

**A blue squares with white text

AI-generated content may be incorrect.**

**Model in use**

A screenshot of a computer screen

AI-generated content may be incorrect.

**03/03/2025**

* Made read me which includes the following:
  + Project Description
  + Data Source
  + Data
  + Processing
  + Data Understanding and Visualisation
  + Algorithm
  + Online Sources
  + Tools and Tech Used

# **K Means Clustering**

**01/03/2025**

* Read up on K means clustering notebook.
* Research dataset to use.
* Dataset link:  <https://grouplens.org/datasets/movielens/>
* Aim of this model is to cluster movies by genres and ratings.

**02/03/2025**

* Implement K Means Clustering

**Cluster distribution by Genres**

A graph of different colored bars

AI-generated content may be incorrect.

**Number of movies in each cluster**

A graph of blue rectangular bars

AI-generated content may be incorrect.

**03/03/2025**

* Made read me which includes the following:
  + Project Description
  + Data Source
  + Data
  + Processing
  + Data Understanding and Visualisation
  + Algorithm
  + Online Sources
  + Tools and Tech Used