Originality report

COURSE NAME
BMCS 2114 (Machine Learning)

STUDENT NAME
YIE SIAN SIT

FILE NAME
G2-G3-Report

REPORT CREATED
Apr 24, 2023

Summary

Flagged passages	21	5%
Cited/quoted passages	1	0.3%

Web matches

kaggle.com	1	2%
support.google.com	4	0.5%
datascienceland.com	2	0.4%
analyticsvidhya.com	2	0.4%
datadiscoverystudio.org	1	0.3%
dblp.org	1	0.3%
opengenus.org	1	0.3%
towardsdatascience.com	1	0.2%
geeksforgeeks.org	1	0.2%
wikipedia.org	1	0.2%
researchgate.net	1	0.2%
stackabuse.com	1	0.2%
archives-ouvertes.fr	1	0.2%
pythonwithliz.com	1	0.2%
linkedin.com	1	0.2%
eumetrain.org	1	0.2%
github.com	1	0.1%

1 of 22 passages
Student passage FLAGGED

In this research, we will show **the** important concepts of **the Bus Breakdown and Delay system** by analyzing and collecting **information from school bus vendors operating out in the field in real time**

Top web match

All information **in the** system is entered by school bus vendor staff.; abstract: **The Bus Breakdown and Delay system** collects **information from school bus vendors operating out in the field in real time**.

Bus Breakdown and Delays. - Data Discovery Studio http://datadiscoverystudio.org/geoportal/rest/metadata/item/f3e1be81f82448e7b64a4e7bf47ace1e/html

2 of 22 passages
Student passage FLAGGED

Bus staff that encounter delays during the route are instructed to radio the dispatcher at the bus vendor's central office. The bus vendor staff are then instructed to log into the Bus Breakdown and…

Top web match

Bus staff that encounter delays during the route are instructed to radio the dispatcher at the bus vendor's central office. The bus vendor staff are then instructed to log into the Bus Breakdown and…

NY Bus Breakdown and Delays | Kaggle https://www.kaggle.com/datasets/new-york-city/ny-bus-breakdown-and-delays

---

3 of 22 passages

Student passage FLAGGED

…arrival and delays in transportation systems, including buses. Different **machine learning algorithms have been** applied in order **to predict delays**

Top web match

machine learning algorithms have been used **to predict** flight **delays** in short-range horizons (e.g., a few hours or days prior to operation).

Distribution Prediction of Strategic Flight Delays via Machine ... https://hal-enac.archives-ouvertes.fr/hal-03873003/document

---

4 of 22 passages

Student passage CITED

Another research showing **a bus arrival time prediction** approach **based on GPS position and real-time traffic flow** was proposed by the authors of the article "A…

Top web match

Dblp: **a bus arrival time prediction** method **based on gps position and real-time traffic flow**. We are hiring!

A Bus Arrival Time Prediction Method Based on GPS Position ... - dblp https://dblp.org/rec/conf/dasc/LeiCLHCZC17

---

5 of 22 passages

Student passage FLAGGED

The Multi-layer **Perceptron (MLP**) algorithm **is a type of artificial neural network that is** commonly **used for** supervised **learning tasks, such as regression**

Top web match

Introduction Multi Layer **Perceptron (MLP) is a type of artificial neural network that is** widely **used for** various machine **learning tasks such as** classification and **regression**.

Multi-Layer Perceptron (MLP): A Basic Understanding - OpenGenus IQ https://iq.opengenus.org/multi-layer-perceptron/

---

6 of 22 passages

Student passage FLAGGED

…using an iterative optimization process called backpropagation, which adjusts **the weights and biases of the** nodes **to minimize the error between the predicted and actual**

Top web match

It helps adjust **the weights and biases of the** network **to minimize the error between the predicted and actual** output.

An Understandable Guide to Backpropagation: From Basics to Mastery https://pythonwithliz.com/an-understandable-guide-to-backpropagation-from-basics-to-mastery/

---

7 of 22 passages

Student passage FLAGGED

Boltzmann machine - a type of neural network that is used for unsupervised learning, by learning a probabilistic model of the input data.

Top web match

Restricted **Boltzmann Machine** (RBM) is **a type of** artificial **neural network that is used for unsupervised learning**. It is a type of generative model that is capable of learning a probability…

Restricted Boltzmann Machine - GeeksforGeeks https://www.geeksforgeeks.org/restricted-boltzmann-machine/

---

8 of 22 passages

Student passage FLAGGED

…memory cells that can selectively store or erase information **over time**, as well as **gates** that **regulate the flow of information into and out of the** cells.

Top web match

The cell remembers values **over** arbitrary **time** intervals and the three **gates regulate the flow of information into and out of the** cell.

Long short-term memory - Wikipedia https://en.wikipedia.org/wiki/Long_short-term_memory

---

9 of 22 passages

Student passage FLAGGED

XGBoost (eXtreme Gradient Boosting) **is a popular machine learning algorithm** that is widely **used for** both **regression** and classification **tasks**. It is based on the concept of gradient boosting…

Top web match

A. Random forest **is a popular Machine learning algorithm used for** classification and **regression tasks** due to its high accuracy, robustness, feature importance, versatility, and scalability.

Random Forest Algorithms - Comprehensive Guide With Examples https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

---

10 of 22 passages

Student passage FLAGGED

…both regression and classification tasks. It is based on **the** concept of **gradient boosting**, which involves iteratively training **an ensemble of weak** prediction **models, such as decision trees, to**

Top web match

... Gradient Boosting: **The Gradient Boosting** algorithm uses **an ensemble of weak models, such as decision trees, to** predict the remaining useful life of ...

Freight Vehicle Travel Time Prediction Using Gradient Boosting ... https://www.researchgate.net/publication/313449746_Freight_Vehicle_Travel_Time_Prediction_Using_Gradient_Boosting_Regression_Tree

---

11 of 22 passages

Student passage FLAGGED

AdaBoost (Adaptive Boosting) **is a machine learning algorithm used for classification and regression tasks**. It is an ensemble learning algorithm that combines the…

Top web match

A. Random forest **is a** popular **Machine learning algorithm used for classification and regression tasks** due to its high accuracy, robustness, feature importance, versatility, and scalability.

Random Forest Algorithms - Comprehensive Guide With Examples https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

---

12 of 22 passages

Student passage FLAGGED

…training a sequence of weak classifiers on a dataset, **with each** classifier **attempting to correct the errors made by the previous ones**. The final prediction is made by taking a weighted…

Top web match

Gradient Boosted Trees (GBT): another ensemble method that builds decision trees sequentially, **with each** tree **attempting to correct the errors made by the previous ones**.

Ali Madani على LinkedIn: #machinelearning #classification ... https://tn.linkedin.com/posts/amlearning_machinelearning-classification-regression-activity-7044325537820901378-vWTR

---

13 of 22 passages

Student passage FLAGGED

The **number of students on the bus at the time of the incident**

Top web match

Data Type: integer Width: - **Number of students on the bus at the time of the incident** as estimated by the staff employed by the reporting bus vendor.

RooLsi/NY-Bus-Breakdown-Delay-Analysis - GitHub https://github.com/RooLsi/NY-Bus-Breakdown-Delay-Analysis

---

14 of 22 passages

Student passage FLAGGED

Count **the number of occurrences of each unique value in** the 'Number_Of_Students_On_The_Bus

Top web match

I want to be able to create a formula that counts **the number of occurrences of each unique value in** that column, with the end goal of graphing it.

Count number of occurrences of each unique occurrence in a list. https://support.google.com/docs/thread/45464746/count-number-of-occurrences-of-each-unique-occurrence-in-a-list?hl=en

---

15 of 22 passages

Student passage FLAGGED

Count **the number of occurrences of each unique value in** the 'Route_Number

Top web match

I want to be able to create a formula that counts **the number of occurrences of each unique value in** that column, with the end goal of graphing it.

Count number of occurrences of each unique occurrence in a list. https://support.google.com/docs/thread/45464746/count-number-of-occurrences-of-each-unique-occurrence-in-a-list?hl=en

---

16 of 22 passages

Student passage FLAGGED

Count **the number of occurrences of each unique value in** the 'Bus_Company_Name

Top web match

I want to be able to create a formula that counts **the number of occurrences of each unique value in** that column, with the end goal of graphing it.

Count number of occurrences of each unique occurrence in a list. https://support.google.com/docs/thread/45464746/count-number-of-occurrences-of-each-unique-occurrence-in-a-list?hl=en

---

17 of 22 passages

Student passage FLAGGED

Count **the number of occurrences of each unique value in** the 'Busbreakdown_ID

Top web match

I want to be able to create a formula that counts **the number of occurrences of each unique value in** that column, with the end goal of graphing it.

Count number of occurrences of each unique occurrence in a list. https://support.google.com/docs/thread/45464746/count-number-of-occurrences-of-each-unique-occurrence-in-a-list?hl=en

---

18 of 22 passages

Student passage FLAGGED

…they have different properties and use cases. MAE is **the average of the** absolute **differences between predicted and actual values**. It measures the magnitude of errors without considering their…

Top web match

The RMSE consists of a metric that calculates **the average of the differences between** the **predicted and actual values** and, then calculates the square root so that the measurement is in the target…

Difference between RMSE and RMSLE - Data Science Blog https://www.datascienceland.com/blog/difference-between-rmse-and-rmsle-656/

---

19 of 22 passages

Student passage FLAGGED

…the absolute differences between predicted and actual values. It **measures the magnitude of errors without considering their direction**. MAE is suitable when you want to have an…

Top web match

The MAE **measures the** average **magnitude of** the **errors** in a set of forecasts, **without considering their direction**. It measures accuracy for continuous variables.

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) https://resources.eumetrain.org/data/4/451/english/msg/ver_cont_var/uos3/uos3_ko1.htm

---

20 of 22 passages

Student passage FLAGGED

…errors are expected to be normally distributed. RMSLE, on **the** other hand, **calculates the** logarithmic **differences between predicted and actual values, and then** takes **the square root** of **the**

Top web match

The RMSE consists of a metric that **calculates** the average of **the differences between** the **predicted and actual values and, then** calculates **the square root** so that **the** measurement is in the target…

Difference between RMSE and RMSLE - Data Science Blog https://www.datascienceland.com/blog/difference-between-rmse-and-rmsle-656/

---

21 of 22 passages

Student passage FLAGGED

Models like **Gradient** Boosted Trees and XGBoost are **ensemble** methods **that** combine **multiple weak models to create a stronger predictive model**

Top web match

Gradient boosting is an **ensemble** method **that** combines **multiple weak models to create a stronger predictive model**. It works by iteratively fitting a new model to the residual errors of the previous…

Gradient Descent vs. Gradient Boosting: A Side-by-Side Comparison https://towardsdatascience.com/gradient-descent-vs-gradient-boosting-a-side-by-side-comparison-7067bb3c5712

---

22 of 22 passages

Student passage FLAGGED

Gradient Boosted Trees **are** ensemble models **that combine** multiple **weak models to create a strong predictive model**. The other models may not be using ensemble techniques…

Top web match

Gradient boosting classifiers **are** a group of machine learning algorithms **that combine** many **weak** learning **models** together **to create a strong predictive model**.

Gradient Boosting Classifiers in Python with Scikit-Learn - Stack Abuse https://stackabuse.com/gradient-boosting-classifiers-in-python-with-scikit-learn/
