

# CODING WEEK-2025

ML\_Task-1

# LEVEL-1:

.The feature\_1 in the task represent the age of the students

- ▶ Justification: The mean of feature\_1 is nearly 17(16.74) and the mean and the mode both are 17 and most data lies between 15 and 17.So,the feature\_1 represents age of students(most probably).
- ▶ The max value of feature 2 is 4 and I plotted heat maps with those whose maximum rating values are 4 I.e. Medu,Fedu,traveltime,failure.The results show that most of the feature 2 is lying for Medu or Fedu greater than 2 OR failures less than 2.So I think the feature 2 represent the effect of education of parents on failures.
- ▶ The max value of Feature 3 is 5 so I plotted heat map with those whose max value is 5 I.e. famrel,freetime,goout,Dalc,health.The results show that most of the feature 3 is strongly related to high values of famrel and freetime ,low values of Dalc.So I think the feature 3 represents time spent with family in free time on a rating of 1-5.

## LEVEL-2:

► There are a total of 10 columns with empty values.

- 1)famsize I have used mode for the famsize because the entries are strings and another reason is I thought mostly there will be a famsize greater than 3(father,mother,2 son's).
- 2)for travel time I used mean.The mode came out to be 1 and the mean came out to be 2(nearly).I thought 30 min is ok to travel from school to home.
- 3)for Fedu I used mean because I thought like say some people will study till Btech some may also interested to do Mtech(for example purpose only) so I thought using mean is good instead of mode and median.
- 4)for higher I used mode.Since the one with no data will do what most students do.
- 5)for freetime I used mode most people have the same free time after school.



6)for absences I used mean because the mode came out to be zero,the mean came out to be around 4.As we don't have data we can not use mode since it is zero so I have used mean.

7)for G2 I used mean.The missed grade will be filled with average grade since the mode of the grades slightly depends on prof.

8)for Feature\_1 I used mode because it its age and most of the students will be of same age i.e. mode of the data.

9)for Feature\_2 I used mode as it is voting based I went to most voted.

10)for Feature\_3 I used mode as it is voting based I went to most voted.

# LEVEL-3:

- 1)schoolsup Vs G3:the results show that most students who are taking schoolsup are getting grades same as most students who are not taking schoolsup.
- 2)G3 Vs Dalc:The results show that those students whose Dalc is more than 3 are getting grades mostly less than 12.
- 3)Internet access Vs G3:The results show that percentage of people for a given grade is almost same for those having internet access and those not having internet access.
- 4)Feature\_1 Vs Romatic:The results show that the students with age 16 and 17 are not likely to have a romantic relationship(mostly).
- 5)G3 Vs famrel:The results show that irrespective of famrel,most people are getting same grades in around 10.

## LEVEL-4:

- ▶ I made a prediction mode for romantic for 3 different data.
  - 1)academic behaviour
  - 2)school and study related
  - 3)life style
- ▶ For each prediction I used 3 classification techniques(used classification because the output will be either yes or no)
  - 1)Logistic Regression
  - 2)Decision Tree
  - 3)Random Forest
- ▶ Also I printed classification report and confusion matrix along with the accuracy which gives the precision of the classifier in each prediction.

# ACCURACY:

<div>data used for prediction</div> <div>Classifier</div>	Academic performance	Life style & Behaviour	School and study related
Logistic Regression	0.615	0.623	0.638
Decision Tree Classifier	0.631	0.546	0.569
Random Forest	0.638	0.562	0.592

# SHAP PLOTS:

- ▶ So for every classifier I have used in prediction model I plotted SHAP plots also. The results show that

## -School and study related vs romantic:

- ~for logistic regression those not studied nursery are most likely to be in romantic relationship.
- ~for Decision Tree the reason for choosing the school is increasing the entropy. So if the reason is course or home then there is increasing chance of having a romantic relationship.
- ~for Random forest the reason for choosing school is highly increasing the entropy and again if the reason is course or home there is more chance of having a romantic relationship.



► Life style and behaviour Vs Romantic:

~for logistic regression the famrel is becoming more important. The one with famrel 3 is having a high chance of romantic relation ship.

~for Decision tree and Random forest the famrel is increasing the entropy mostly. The predictions say that having high famrel is very likely to have a romantic relation ship.

► Academic performance Vs Romantic:

~for logistic regression a G3 value of 19 and a G1 value of 17 are slightly having more chance of having a romantic relation ship but a G2 value of 18 is decreasing the chance of having romantic relation ship very drastically.

~for Decision tree and Random Forest the G2 increasing the entropy mostly. A low value of G2 is very unlikely to have a romantic relation ship. but on the other way having a high G2 is very likely to have a romantic relation ship.