PROJECT

Student Flexibility in Online Learning

ABOUT THE DATASET

The dataset provides a comprehensive view of student flexibility in online learning across various parameters.

It includes features such as educational level, institution type, gender, age, device used for accessing online resources, IT student status, location, financial condition, internet type, and network type.

Encompasses a diverse range of students, from primary school to university, residing in different locations with varying financial backgrounds.

Factors like internet connectivity, device accessibility, and financial stability are key determinants of students' flexibility in engaging with online learning.

Analysis of this dataset offers valuable insights into enhancing online learning experiences and ensuring inclusivity among student populations.

Data Info - Orange

Data table properties

Name: dm

Size: 1205 rows, 11 columns

Features: 9 categorical, 1 numeric

Targets: categorical outcome with 3 classes

Additional attributes

1205

Data Table (1) - Orange

Info

1205 instances (no missing data)
10 features
Target with 3 values
No meta attributes.

Variables

- [x] Show variable labels (if present)
- [ ] Visualize numeric values
- [x] Color by instance classes

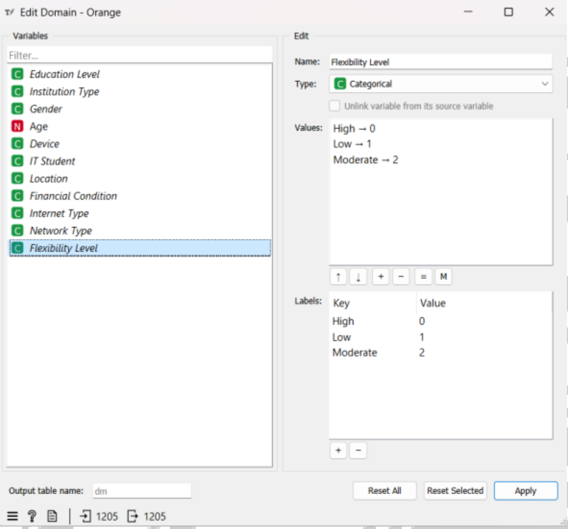
Selection

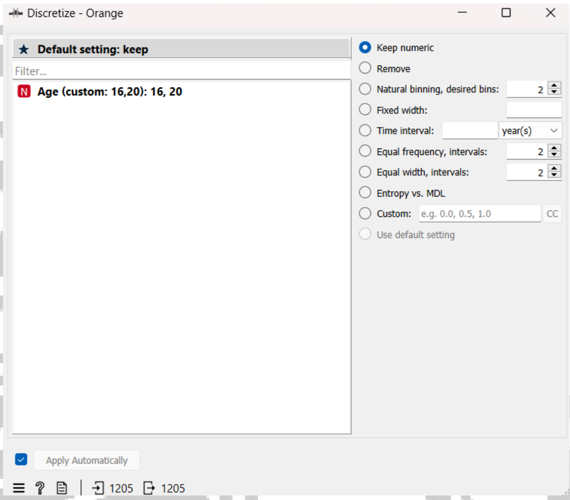
- [x] Select full rows

	Flexibility Level	Education Level	Institution Type	Gender	Age	Device	IT Student	Location	Financial Condition	Internet Type	Network Type
1	Moderate	University	Private	Male	23	Tab	No	Town	Mid	Wifi	4G
2	Moderate	University	Private	Female	23	Mobile	No	Town	Mid	Mobile Data	4G
3	Moderate	College	Public	Female	18	Mobile	No	Town	Mid	Wifi	4G
4	Moderate	School	Private	Female	11	Mobile	No	Town	Mid	Mobile Data	4G
5	Low	School	Private	Female	18	Mobile	No	Town	Poor	Mobile Data	3G
6	Low	School	Private	Male	11	Mobile	No	Town	Poor	Mobile Data	3G
7	Low	School	Private	Male	11	Mobile	No	Town	Mid	Wifi	4G
8	Moderate	School	Private	Male	11	Mobile	No	Town	Mid	Wifi	4G
9	Low	College	Public	Male	18	Mobile	No	Town	Mid	Wifi	4G
10	Moderate	School	Private	Male	11	Mobile	No	Town	Mid	Mobile Data	3G
11	Low	University	Public	Female	18	Mobile	No	Town	Mid	Wifi	4G
12	Low	College	Private	Female	18	Mobile	No	Town	Mid	Wifi	4G
13	Moderate	School	Private	Male	11	Mobile	Yes	Town	Mid	Mobile Data	3G
14	Low	College	Private	Female	18	Mobile	No	Rural	Mid	Wifi	4G
15	Moderate	School	Private	Female	11	Mobile	No	Town	Poor	Mobile Data	3G
16	Low	School	Private	Male	11	Mobile	No	Town	Poor	Wifi	4G
17	Low	University	Public	Female	27	Computer	Yes	Town	Poor	Mobile Data	4G
18	Low	University	Public	Female	23	Mobile	No	Rural	Mid	Wifi	4G
19	Low	College	Private	Female	18	Mobile	No	Town	Mid	Wifi	4G
20	Moderate	College	Private	Female	18	Mobile	No	Town	Mid	Wifi	4G
21	High	School	Private	Female	10	Mobile	No	Town	Rich	Wifi	4G
22	Low	School	Public	Male	10	Mobile	No	Town	Mid	Mobile Data	4G
23	Low	College	Private	Female	18	Mobile	No	Rural	Mid	Mobile Data	3G
24	Moderate	School	Private	Female	11	Mobile	No	Town	Mid	Mobile Data	3G
25	Moderate	College	Public	Male	18	Mobile	No	Town	Mid	Wifi	4G
26	Moderate	School	Private	Male	11	Mobile	No	Town	Mid	Mobile Data	4G
27	Moderate	School	Private	Female	9	Mobile	No	Town	Mid	Mobile Data	4G
28	Moderate	University	Private	Male	27	Computer	Yes	Town	Mid	Wifi	4G
29	Moderate	School	Private	Female	9	Mobile	No	Town	Mid	Mobile Data	4G
30	Moderate	College	Private	Female	18	Mobile	No	Town	Mid	Wifi	4G
31	Moderate	School	Private	Female	10	Mobile	No	Town	Mid	Mobile Data	3G
32	Low	College	Private	Female	18	Mobile	No	Rural	Mid	Mobile Data	4G
33	Moderate	University	Private	Male	23	Mobile	Yes	Town	Mid	Mobile Data	4G

PREPROCESSING

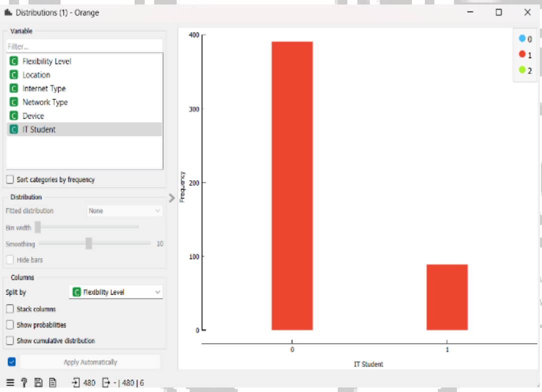
Categorical to Numerical Encoding: Categorical variables such as educational level, institution type, gender, location, financial condition, internet type, and network type were encoded into numerical values. This step facilitates machine learning algorithms to work effectively with categorical data.Age Discretization by Binning: Age values were discretized by binning them into distinct age groups. This process transforms continuous age values into categorical bins, making it easier to analyze age-related trends without losing significant information.



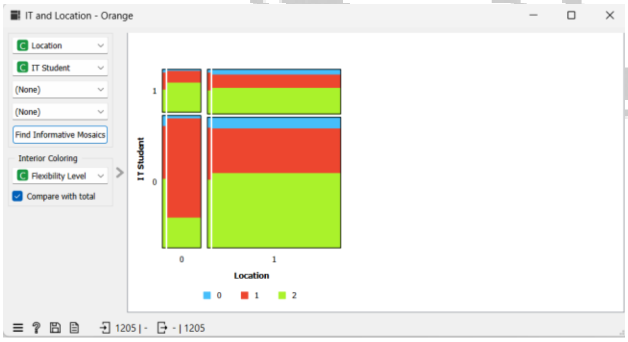


DATA VISUALIZATION

Distribution on IT Students and Flexibility Level



Mosaic Display on IT Student,Location and the Flexiblity level



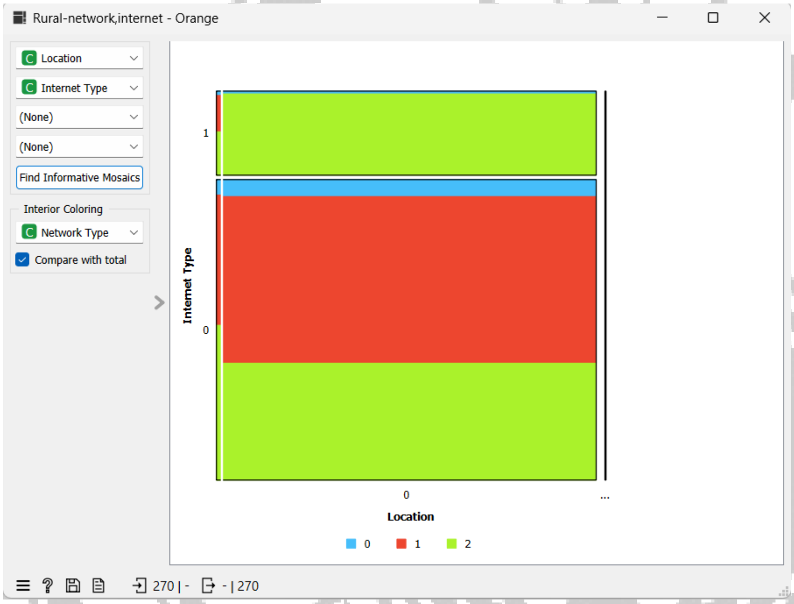
After conducting a mosaic display on IT students, location, and flexibility level in the dataset, we observed that students opting for IT studies are distributed fairly evenly between village and urban areas, with percentages of approximately 24.04% and 23.03% respectively. This suggests that location alone may not be a significant determining factor in students' decisions to pursue IT fields.

From the mosaic display analysis, we discovered a trend indicating that flexibility levels are notably lower among students residing in rural areas who have not chosen IT fields. This suggests that there might be a correlation between lower flexibility levels and rural residence among students who opt out of IT studies.

However, it's crucial to note that this observation doesn't establish a significant relationship between choosing the IT field and location alone. Rather, it hints at a potential link between lower flexibility levels and rural environments.

By aggregating the data of students with low flexibility levels, particularly those residing in rural areas, we can infer that flexibility levels tend to be lower in rural settings.

Mosaic Display on Location,Internet Type,Network Type in Rural Area

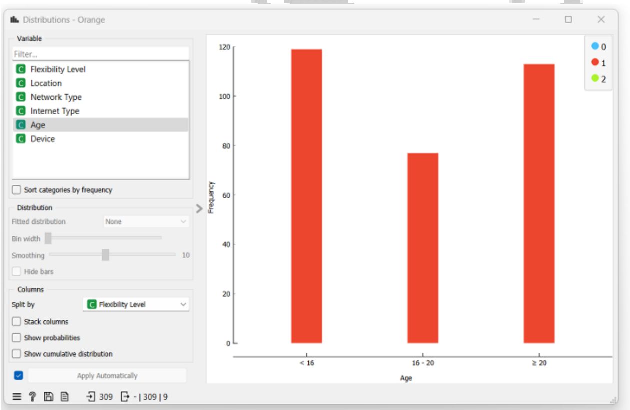


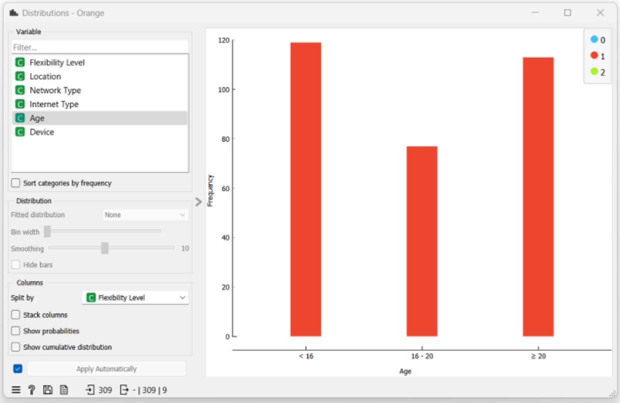
Conclusion

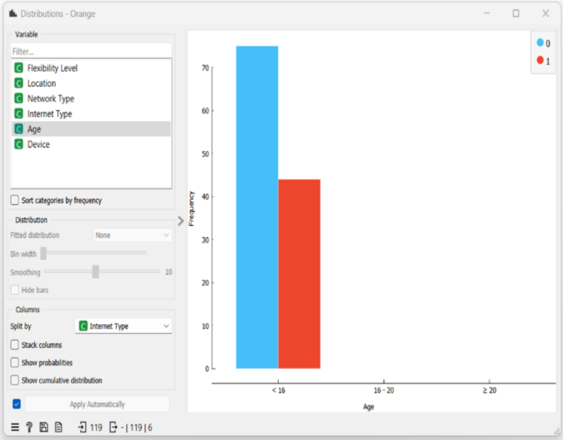
Upon further analysis focusing solely on rural areas, we discovered a predominant usage of mobile phones, particularly with 3G internet and mobile data. This finding suggests a potential reason for the observed low flexibility levels, as 3G speed and mobile data may not adequately support prolonged online classes.

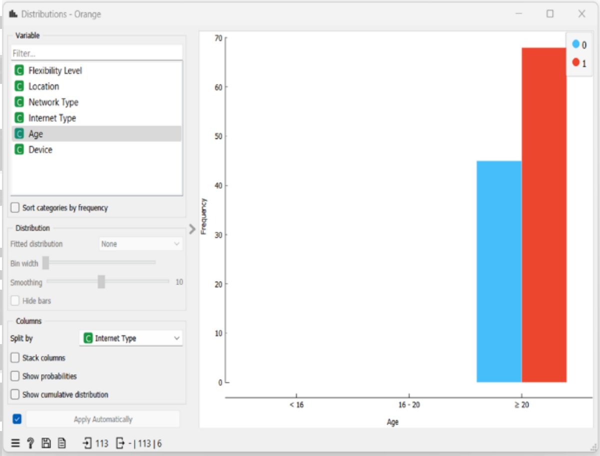
The prevalence of mobile devices with limited internet capabilities in rural settings could pose significant challenges for students in accessing and engaging with online learning resources effectively.

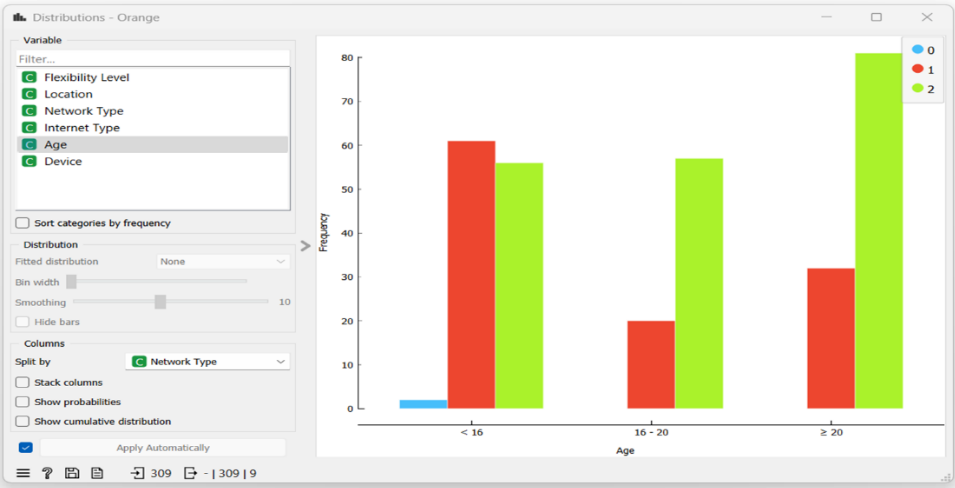
HISTOGRAM ON AGE AND FLEXIBILITY LEVEL IN URBAN AREA

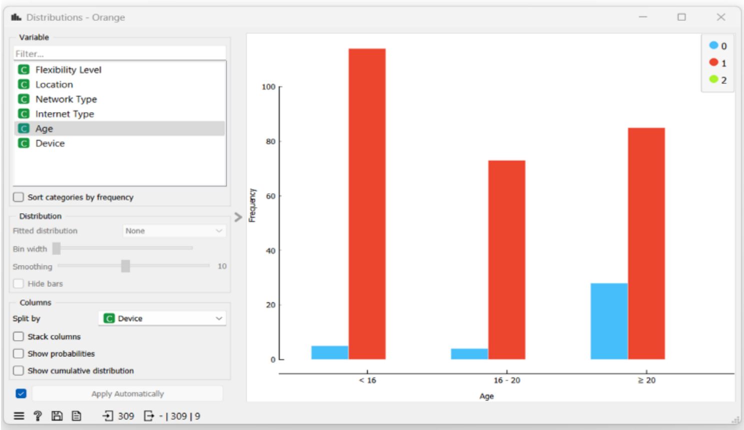



Distributions - Orange
Variable
Flexibility Level
Location
Network Type
Internet Type
Age
Device
Sort categories by frequency
Distribution
Fitted distribution
None
Bin width
Smoothing
Hide bars
Columns
Split by
Flexibility Level
Stack columns
Show probabilities
Show cumulative distribution
Apply Automatically
309
- | 309 | 9
Frequency
Age
< 16
16 - 20
≥ 20
0
1
2


Distributions - Orange
Variable
Flexibility Level
Location
Network Type
Internet Type
Age
Device
Split by
Internet Type
Apply Automatically
119
- | 119 | 6
Frequency
Age
< 16
16 - 20
≥ 20
0
1


Distributions - Orange
Variable
Flexibility Level
Location
Network Type
Internet Type
Age
Device
Split by
Internet Type
Apply Automatically
113
- | 113 | 6
Frequency
Age
< 16
16 - 20
≥ 20
0
1


Distributions - Orange
Variable
Filter...
Flexibility Level
Location
Network Type
Internet Type
Age
Device
Sort categories by frequency
Distribution
Fitted distribution
None
Bin width
Smoothing
10
Hide bars
Columns
Split by
Network Type
Stack columns
Show probabilities
Show cumulative distribution
Apply Automatically
309
- | 309 | 9
Frequency
Age
< 16
16 - 20
≥ 20
0
1
2


Distributions - Orange
Variable
Flexibility Level
Location
Network Type
Internet Type
Age
Device
Split by
Device
Apply Automatically
309
- | 309 | 9
Frequency
Age
< 16
16 - 20
≥ 20
0
1
2

Conclusion

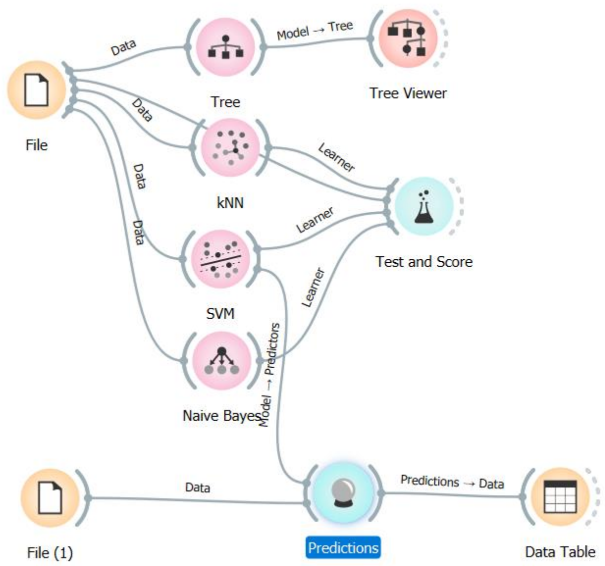
•In our Urban area students within age<16 and >20 are mostly having flexibility low.

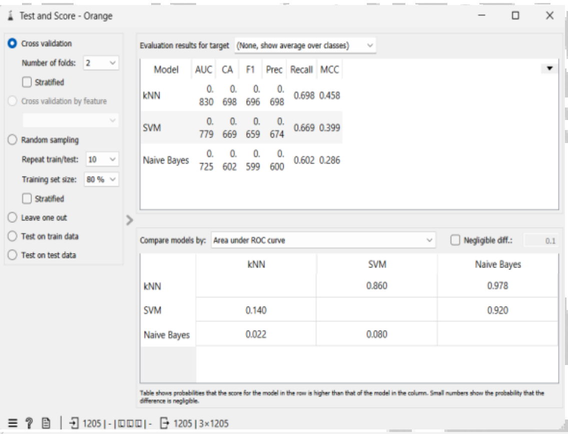
•Students within age<16 are mostly using mobile data and students within age>20 are mostly using wifi. So internet type is not affecting.

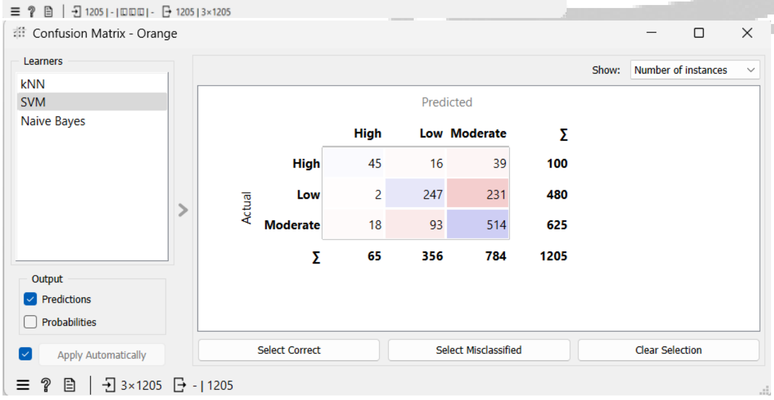
•Students within age<16 are using both 3G and 4G and students within age>20 are mostly using 4G.So students have adequate network access therefore network type is also not affecting their flexibility.

•Only factor common between them is their device usage i.e mobile. So mobile might be reason for the low flexibility in urban region.

DATA MODELLING







Predictions - Orange

Show probabilities for (None)

	SVM	Education Level	Institution Type	Gender	Age	Device	IT Student	Location	inancial Conditio	Internet Type	Network Type
1	Low	College	Private	Female	18	Mobile	No	Rural	Mid	Mobile Data	3G

After conducting various data modeling techniques including Support Vector Machine (SVM), Decision Tree, Naive Bayes, and k-Nearest Neighbors (KNN), we found that the SVM model demonstrated the highest accuracy among them. Utilizing the SVM model, we proceeded to predict a new dataset and obtained the classification results.

The SVM model's superior performance underscores its effectiveness in accurately classifying the data based on the provided features. Leveraging this model, we achieved reliable predictions for the new dataset, which can offer valuable insights into the classification of similar data instances.