greatlearning
Learning for Life

GREAT LAKES
INSTITUTE OF MANAGEMENT
Global Mindset - Indian Roots

Group-5

Capstone Project - Interim Report

Motion Capture Hand Postures Recognition

Industry Review:

- Industry Review: Technological developments are the capstone of today's era, but even after decades of evolution and a bright foreseeable time ahead this technology still treats the human body system as its ideal. Consider our heart, the best pumps in the world still cannot match the human heart's efficiency, same goes with other organs be it lungs, kidneys or any of it. Our hands are no different, the smartest robots are trying to replicate the movement and the agility of the human hand. So why not utilize what we have already, combining it with the technology we possess and try to gain maximum from it.

There can be numerous hand postures, but even considering a handful of them and utilizing them can be very beneficial. Its implementation is not something which is new, I guess we all have seen news channels trying to communicate with the hearing impaired via hand gestures (sign language). Coming to today's scenario, its implementations have taken a gigantic boost. These are now used in computer human interaction, video games, 3D animations, movies, etc. etc. and this can be said with confidence that this is just the start.

The future holds a lot for this field as well, multiple researches in this field are going on which shows high potential. It has showed tremendous results in robotics, a whole domain in neural network is now present for researchers to advance in the field. Virtual world experience is yet another revolutionizing example.

- Literature Survey: There have been multiple papers published in this segment. Some of them are by: A. Gardner, J. Kanno, C.A. Duncan and R. Selmic on 'Measuring distance between unordered sets of different sizes,' in 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014, pp. 137-143. Another one is again by A. Gardner, C. A. Duncan, J. Kanno, and R.

Selmic on '3D hand posture recognition from small unlabelled point sets,' in 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC), Oct 2014, pp. 164-169.

There applications as discussed earlier is also vast which includes human computer interaction, video gaming, virtual environments, 3D models, movies, traffic regulation, aeronautics, etc.

Dataset and Domain:

- Data Dictionary:

- 'Class': The class ID of the given record. Ranges from 1 to 5 with 1=Fist(with thumb out), 2=Stop(hand flat), 3=Point1(point with pointer finger), 4=Point2(point with pointer and middle fingers), 5=Grab(fingers curled as if to grab).
'User' - Integer. The ID of the user that contributed the record. No meaning other than as an identifier.

- 'Xi': The x-coordinate of the i-th unlabelled marker position. 'i' ranges from 0 to 11.

- 'Yi': The y-coordinate of the i-th unlabelled marker position. 'i' ranges from 0 to 11.

- 'Zi': The z-coordinate of the i-th unlabelled marker position. 'i' ranges from 0 to 11.

Each record is a set. The i-th marker of a given record does not necessarily correspond to the i-th marker of a different record.

- Variable Categorization:

- Numeric: 36 variables

- Categorical: 2 variables

- Pre-Processing Data Analysis:

Count of missing values-

Class	0		User	0	
X0	0	Y0	0	Z0	0
X1	0	Y1	0	Z1	0
X2	0	Y2	0	Z2	0
X3	690	Y3	690	Z3	690

X4	3120	Y4	3120	Z4	3120
X5	13023	Y5	13023	Z5	13023
X6	25848	Y6	25848	Z6	25848
X7	39152	Y7	39152	Z7	39152
X8	47532	Y8	47532	Z8	47532
X9	54128	Y9	54128	Z9	54128
X10	63343	Y10	63343	Z10	63343
X11	78064	Y11	78064	Z11	78064

- Project Justification:

Problem statement:

Hand gestures are natural, intuitive and require almost no learning or remembering. It has a vast implementation specially in this technology centric era.

So, working on a very small section of this gigantic domain we recognize and try to predict the hand postures by analyzing the co-ordinates of multiple markers placed on the left hand. The prediction has to be done out of 5 postures via classification method by analyzing the movement of the markers.

Project Outcome:

Commercial: To create a reliable model by converting raw coordinate data into machine understandable hand postures. These postures can be further used to create other meaningful data such as written complex sentences for the hearing impaired.

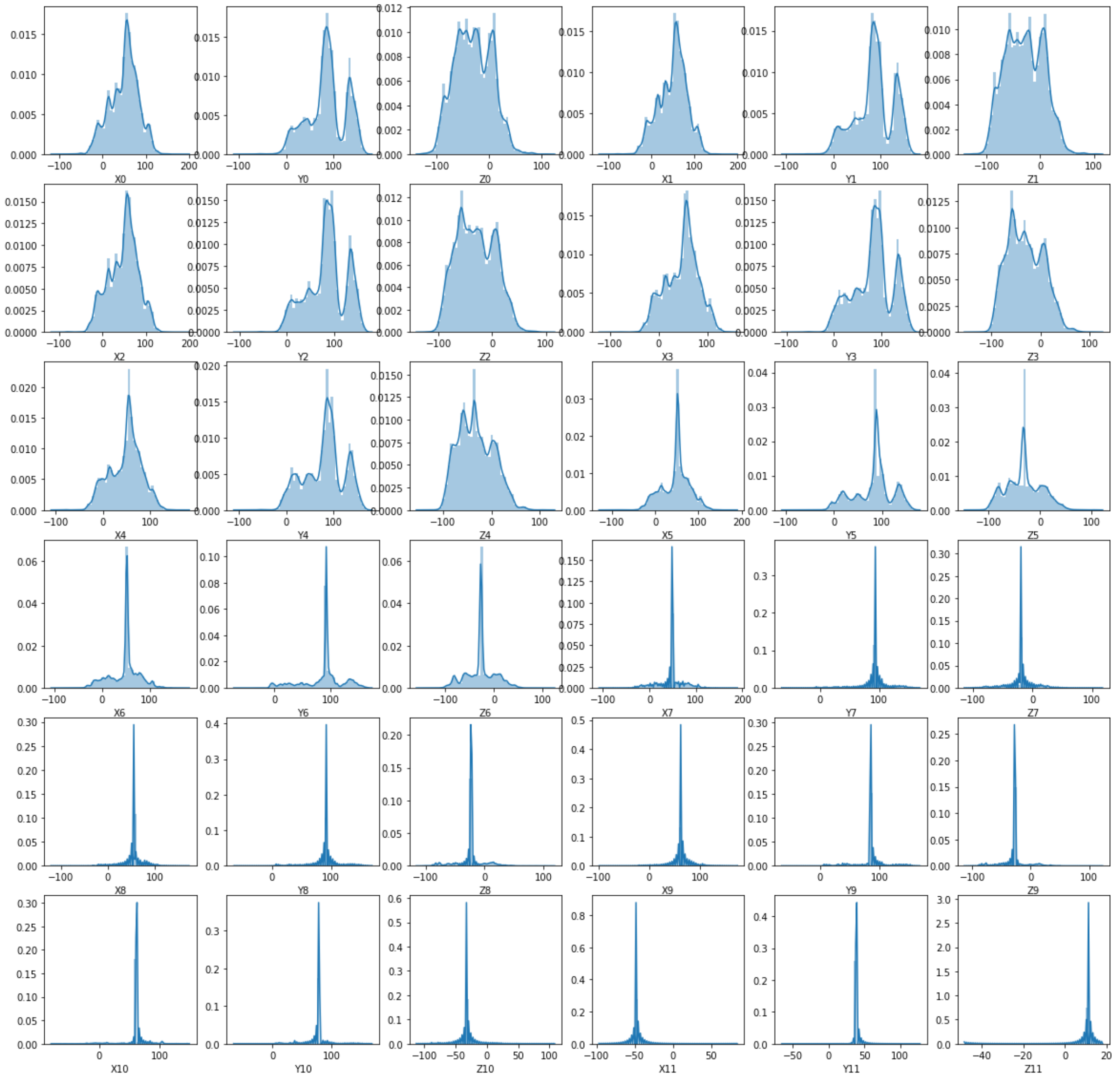
Academic: This project has a great scope in terms of research and development. There are a still numerous fields where they are being developed and implemented like robotics, aeronautics.

Social: To maximize the accuracy of the correct recreation of a physical hand signal into the digital world. The real-world hand movements translated in the digital language have various implementations.

Data Exploration (EDA):

- <u>Distribution of variables</u>:

Univariate analysis:



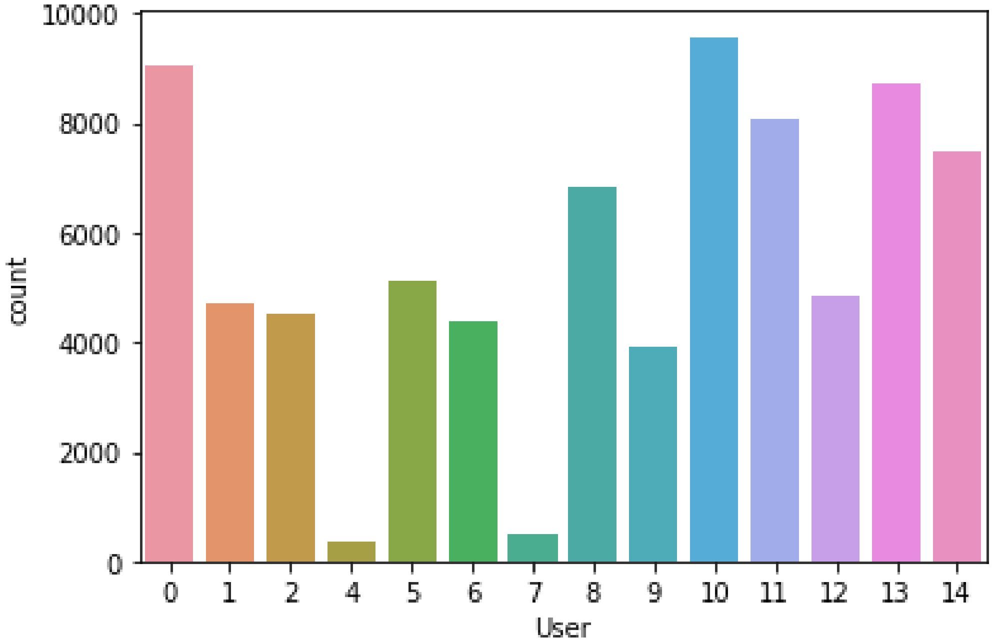
<u>Inferences</u>:

- There are a lot of features towards the second half of the dataset which are concentrated near one value, the reason for this is the imputation we had used.

- As these features had most of their data missing, we had imputed median values of their features at their places resulting in a distribution concentrated near one value.

Users:

The split has to be on the basis of every user, keeping one user as test and others as train, repeating the process for all users. So, this feature is very important.



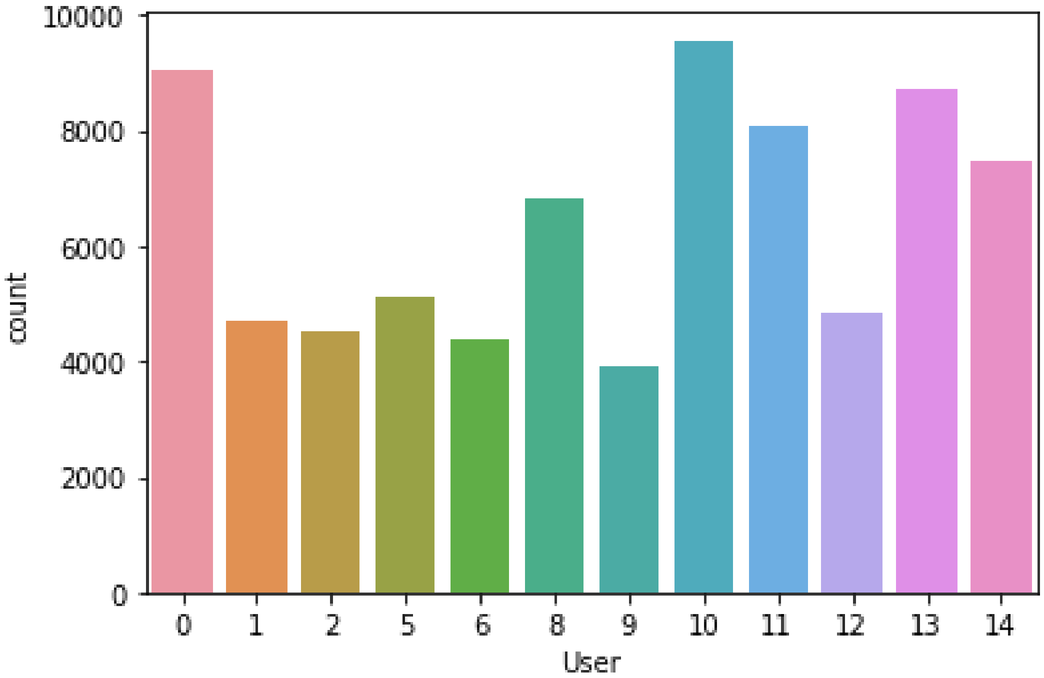
<u>Inferences:</u>

- We have users 0 to 14 but user number 3 is missing from the list, so we have a total of 14 users.

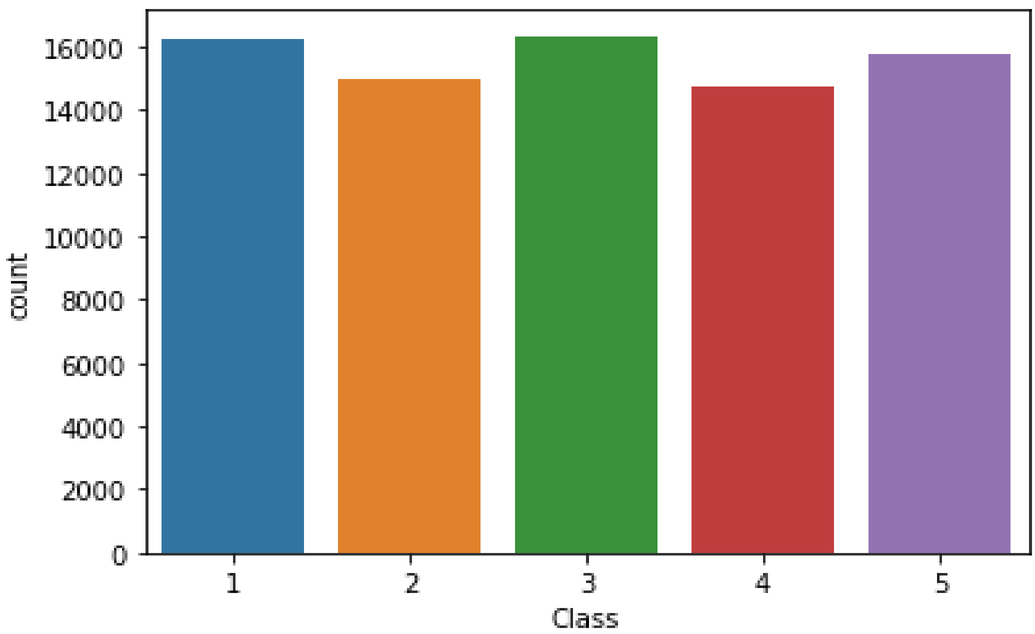
- There is severe imbalance when it comes to users, as this is not the target feature, we cannot apply advanced features like SMOTE to rectify the imbalance but still needs to be treated.

- The user number 4 and 7 have very less values as compared to others. So, it will be better to drop these users to avoid discrepancies later.

- After dropping both the users, we have a total of 12 users.



Target Column:



<u>Inference:</u>

- There is no imbalance seen in the target column (Class) after dropping the 0^{th} User.

- The 5 poses are not equal but they are near to each other exhibiting a real-world problem scenario.

- Null Value Treatment:

As seen earlier most of the features had missing values in the form of '?'. Some of the features had them in abundance up to 99.9% missing values.

Class	0.000000
User	0.000000
X0	0.000000
Y0	0.000000
Z0	0.000000
X1	0.000000
Y1	0.000000
Z1	0.000000
X2	0.000000
Y2	0.000000
Z2	0.000000
X3	0.883539
Y3	0.883539
Z3	0.883539
X4	3.995134
Y4	3.995134
Z4	3.995134
X5	16.675844
Y5	16.675844
Z5	16.675844
X6	33.098150
Y6	33.098150
Z6	33.098150
X7	50.133811
Y7	50.133811
Z7	50.133811
X8	60.864332
Y8	60.864332
Z8	60.864332
X9	69.310455
Y9	69.310455
Z9	69.310455
X10	81.110186
Y10	81.110186
Z10	81.110186
X11	99.960305
Y11	99.960305
Z11	99.960305

Treatment:

So, first the features which had missing values greater than 65% of the total values were removed. The results were checked and then in the next iteration the features which had 60% of the values missing were removed and the results were checked.

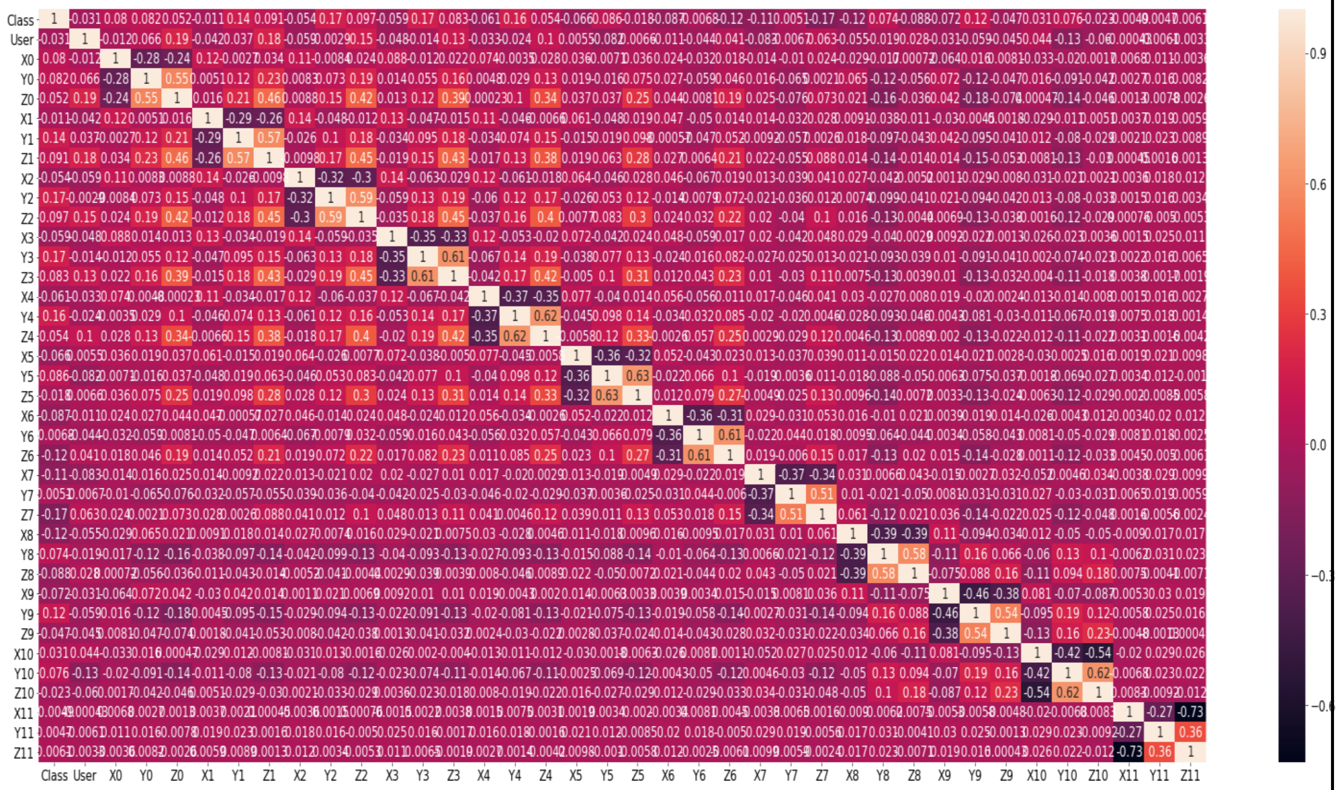
This iterative process was done again and again to see which values were impacting the performance of the model.

Finally, the features from X0 to Z8 were kept, and the others were dropped.

The remaining features with missing values were first converted to null values and then treated by performing median imputation for every feature they belonged.

- Check for Multicollinearity:

Heatmap: Heatmap comprise of the correlation matrix which demonstrates the correlation between all the features among themselves.



<u>Inferences:</u>

- The target variable does not have a high correlation with any of the features.

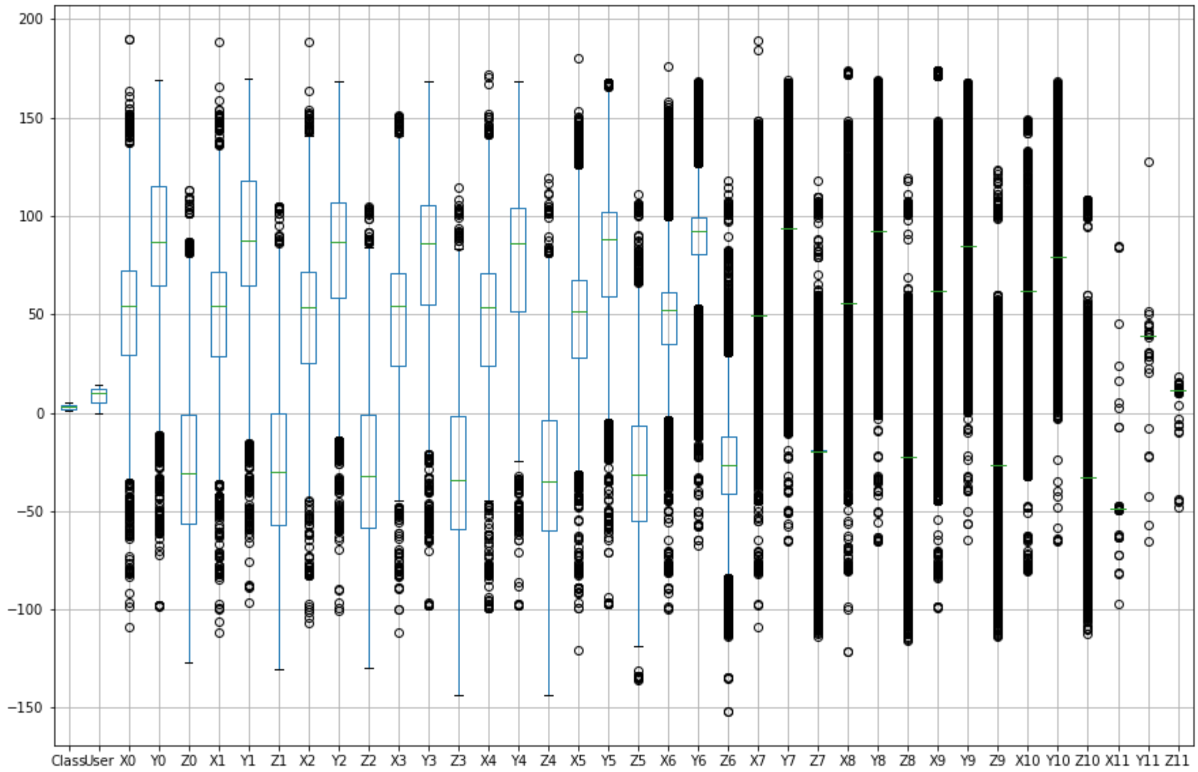
- It can be seen that there are a few correlations among features like X11 and Z11, also X4 and Y4.

- These correlations are probably because of the missing value imputations.

- Another probable reason could be that as they are co-ordinates, their positions can overlap or be carried out in a certain pattern similar to others thus causing the correlation.

- Outliers and their treatment:

Boxplot:



Inferences:

- The circles represent the values which are considered as outliers.

- It can be seen that all the numerical features have numerous outliers.

- Also, these outliers constitute a major portion of the dataset and removing them will lead to a severe data loss.

Treatment:

Instead of removing these outliers we will cap them according to their classes.

We first divide the dataset according to the 5 unique classes and then the values in the features above (Q3+1.5*IQR) are capped as Q3 similarly for values below (Q1-1.5*IQR) are capped as Q1.

This caps the outliers according to the classes, finally we merge all the classes to make the dataset full again.

This will not remove the outliers in the whole data but treat them to give us better results.

- Statistical Significance of variables:

	count	mean	std	min	25%	50%	75%	max
Class	77225.0	2.985147	1.424904	0.000000	2.000000	3.000000	4.000000	5.000000
User	77225.0	7.984668	4.715390	0.000000	5.000000	10.000000	12.000000	14.000000
X0	77225.0	50.170997	32.690938	-108.552738	29.182794	54.447231	72.337860	190.017835
Y0	77225.0	86.295927	39.922233	-98.233756	64.469911	86.862293	115.132081	169.175464
Z0	77225.0	-29.901672	34.436921	-126.770872	-56.384867	-30.714133	-1.235866	113.345119
X1	77225.0	49.459413	32.471890	-111.685241	28.588047	54.094346	71.611951	188.691997
Y1	77225.0	86.670726	40.176686	-96.142589	64.946287	87.842006	118.343028	170.209350
Z1	77225.0	-29.385080	34.812377	-130.062866	-57.339418	-29.966795	-0.197978	104.697852
X2	77225.0	48.530891	33.616178	-106.886524	25.025493	53.758460	71.474641	188.760168
Y2	77225.0	84.132895	40.863843	-100.789312	58.690936	86.733377	107.051325	168.186466
Z2	77225.0	-30.440980	35.160349	-129.595296	-58.610455	-32.175755	-0.756984	104.590879
X3	76559.0	48.393729	33.895571	-111.761053	23.667194	54.064771	71.326982	151.033472
Y3	76559.0	82.335346	41.464138	-97.603414	54.396429	85.932655	105.662474	168.292018
Z3	76559.0	-31.044171	35.976243	-143.540529	-59.356624	-33.836915	-1.214481	114.624261
X4	74224.0	48.349916	34.230702	-99.107635	22.778466	53.791478	71.916535	172.275978
Y4	74224.0	80.656584	42.491489	-97.948829	49.215670	85.897709	105.539848	168.258643
Z4	74224.0	-31.933970	36.443741	-143.282741	-60.550543	-35.073132	-2.020195	119.237203
X5	64491.0	47.009420	34.839667	-120.657868	19.382493	51.919943	71.885554	180.563322
Y5	64491.0	81.543759	42.809730	-97.468548	49.295779	88.284443	107.930600	167.926171
Z5	64491.0	-30.257636	36.896156	-135.699430	-59.041017	-31.695786	0.200423	110.898899
X6	51839.0	45.641206	36.360616	-100.084275	15.430550	52.043732	72.588132	176.409004
Y6	51839.0	83.858566	43.052158	-67.283707	51.818881	92.238074	111.446099	168.598384
Z6	51839.0	-26.619605	35.706090	-151.838668	-55.384731	-26.389513	2.481813	117.914907
X7	38719.0	44.414635	38.401930	-108.605639	13.021478	49.489645	75.700520	189.221529
Y7	38719.0	88.577974	40.416270	-64.972157	63.591562	93.828092	119.358133	169.127359
Z7	38719.0	-20.354611	33.927060	-113.733105	-45.646516	-19.380642	6.607342	117.815967
X8	30455.0	48.235310	38.603879	-121.182089	17.068893	55.990979	79.649726	173.906643
Y8	30455.0	86.074982	41.570010	-65.077550	53.681984	92.201766	121.947521	169.322843
Z8	30455.0	-24.378889	35.738127	-115.951733	-52.561040	-22.497355	6.410337	119.213101

Inferences:

- There are 77225 distinct observations which include the co-ordinates of 12 different users with 5 different hand gestures.

- The values of the co-ordinates range between -151 to 190.

- The median values range between -35 to 93.

Machine Learning:

- Evaluation on Leave-One-User out basis:

The problem statement demanded that the 12 users be divided into train and test one by one, where every iteration has one user as test set and the other users as train sets. So that the evaluation can be done on the basis of each user, how each user is performing.

This can be done simply by running a loop for the unique users in the dataset, iterating again and again keeping one user as test set and all others as training set. The same is done for the target variable also.

- Base Model:

We have used Logistic Regression as our base model, implementing on the loop we just defined above evaluating the model on leave-one-user out basis.

The above described loop will have Logistic Regression defined as a variable say “lr” and then will be fit on the training set and predicting on the test set for every iteration.

This process will repeat for all the 12 users, and at the end an evaluation of the desired score will be done.

We also printed out the confusion matrix for every user, to see how they perform individually.

The base model clearly under performs in this scenario, but the main objective of applying a base model is to get a rough intuition of how our dataset is behaving.

These base models be it Logistic Regression or Decision Tree or any other model are considered as weak learners. So, our next step would be to combine these weak learners (base model) using ensemble techniques to improve on our model’s performance.

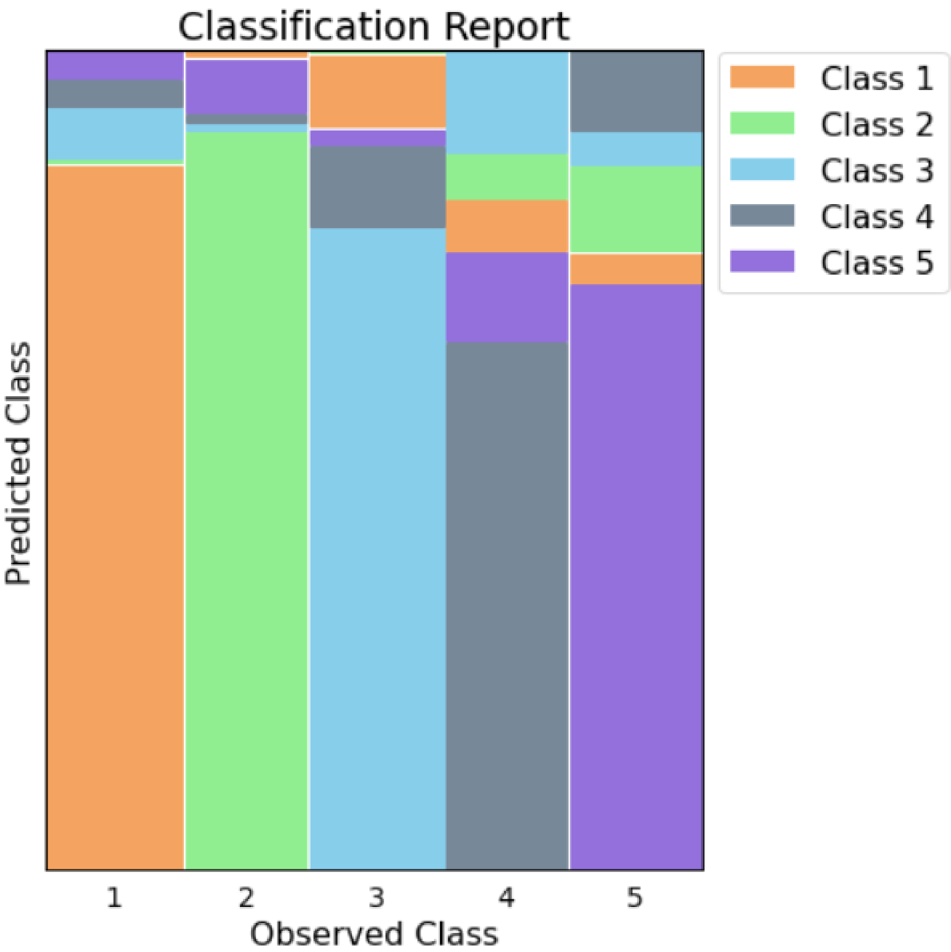
- F1 Score and the Accuracy Score:

As there is almost negligible imbalance in the dataset the F1 score and the accuracy score will give approximate results:

```
The F1 score for the base model is 0.6435747299399242
The Accuracy score for the base model is 0.6649818248819156
```

- Result Visualization:

The result is visualized using a defined function mosaic plot which creates mosaic plots to visualize the results. This is the mosaic plot for the base model using Logistic Regression.



- Confusion Matrix:

The confusion matrix is made for every user separately just to show how are performing in each of the 5 classes.

```
[[4278, 11, 303, 151, 166],
 [31, 3973, 27, 39, 284],
 [425, 10, 3827, 479, 83],
 [260, 235, 538, 2812, 463],
 [164, 500, 186, 464, 3459]]
```

Inferences:

- This confusion matrix represents just one iteration of the model i.e. for just one user.

- Here, the diagonals represent the true values of all each class.

- The non-diagonal values are the ones predicted wrong.

- For instance, the last value of the 1st row is 166, it represents that actually that value belonged to Class 5 but it was predicted as Class 1. Similarly, for other non-diagonal values.

- The lower the non-diagonal values the higher will be the diagonal values and higher will be the performance of the model for that particular user.

- The exact same thing is represented by the mosaic plot shown above.

- One thing is clear after observing the mosaic plot, that is the base model is under performing and we need ensemble approach to get better results.

Ensemble Techniques:

- Random Forest:

This special bagging technique improved the performance of the model by quiet a lot.

The F1 and the accuracy scores jumped up by a considerable margin. The implementation was same as done in the case of base model.

Scores:

```
The F1 score for the Random Forest model is 0.89825
The Accuracy score for the Random Forest model is 0.90825
```

- XGBoost:

This boosting technique showed results similar to that of Random Forest.

Scores:

```
The F1 score for the Random Forest model is 0.89087
The Accuracy score for the Random Forest model is 0.90075
```

- Support Vector Machine:

This advanced algorithm improved the performance of the model.

The F1 and the accuracy scores jumped up a little.

Scores:

```
The F1 score for the Support Vector Machine algorithm is: 0.9179
The Accuracy score for the Support Vector Machine algorithm is: 0.9223
```

- Voting Classifier:

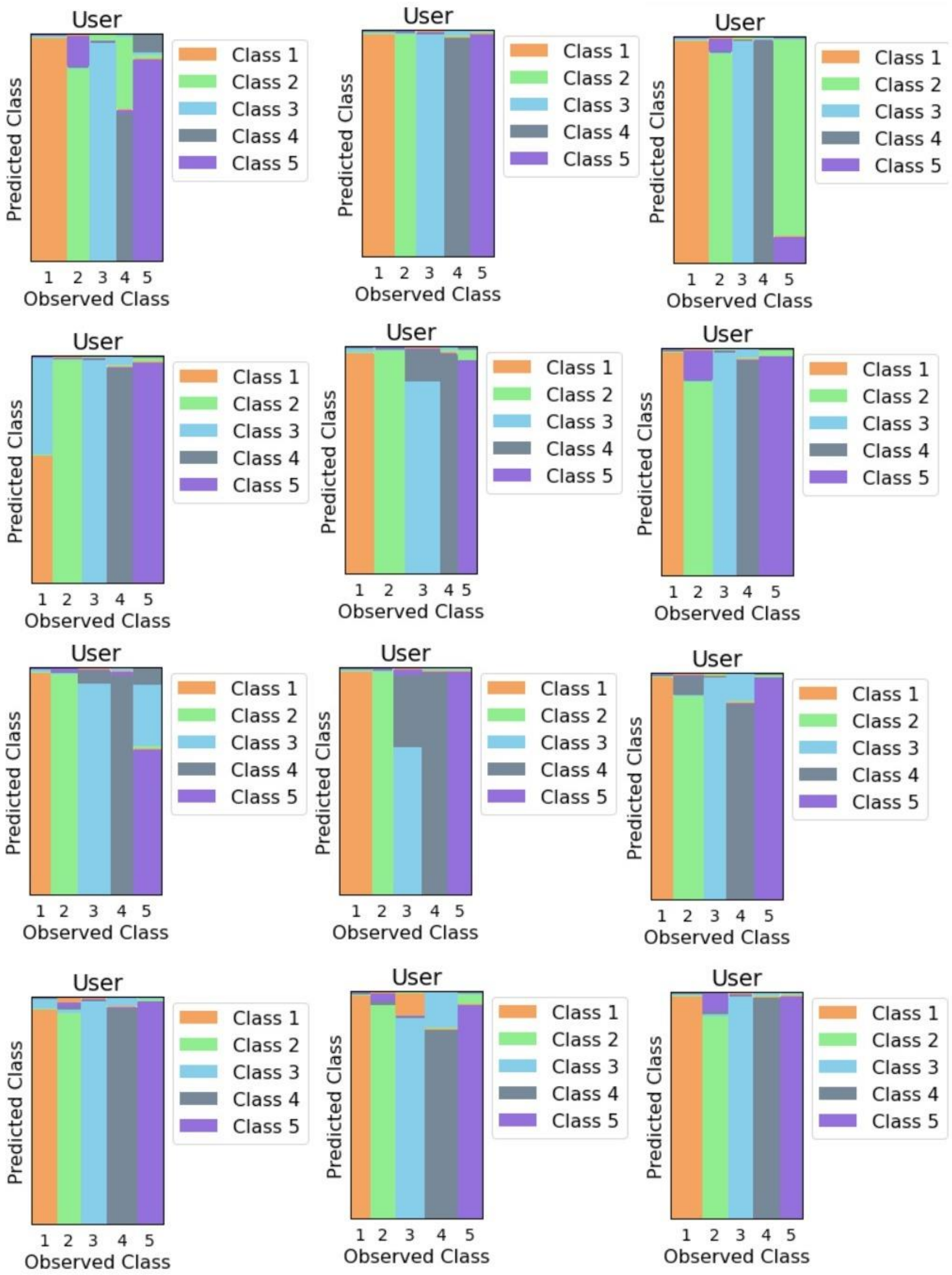
This special bagging technique improved the performance of the model by quiet a lot.

The F1 and the accuracy scores jumped up a little.

Scores:

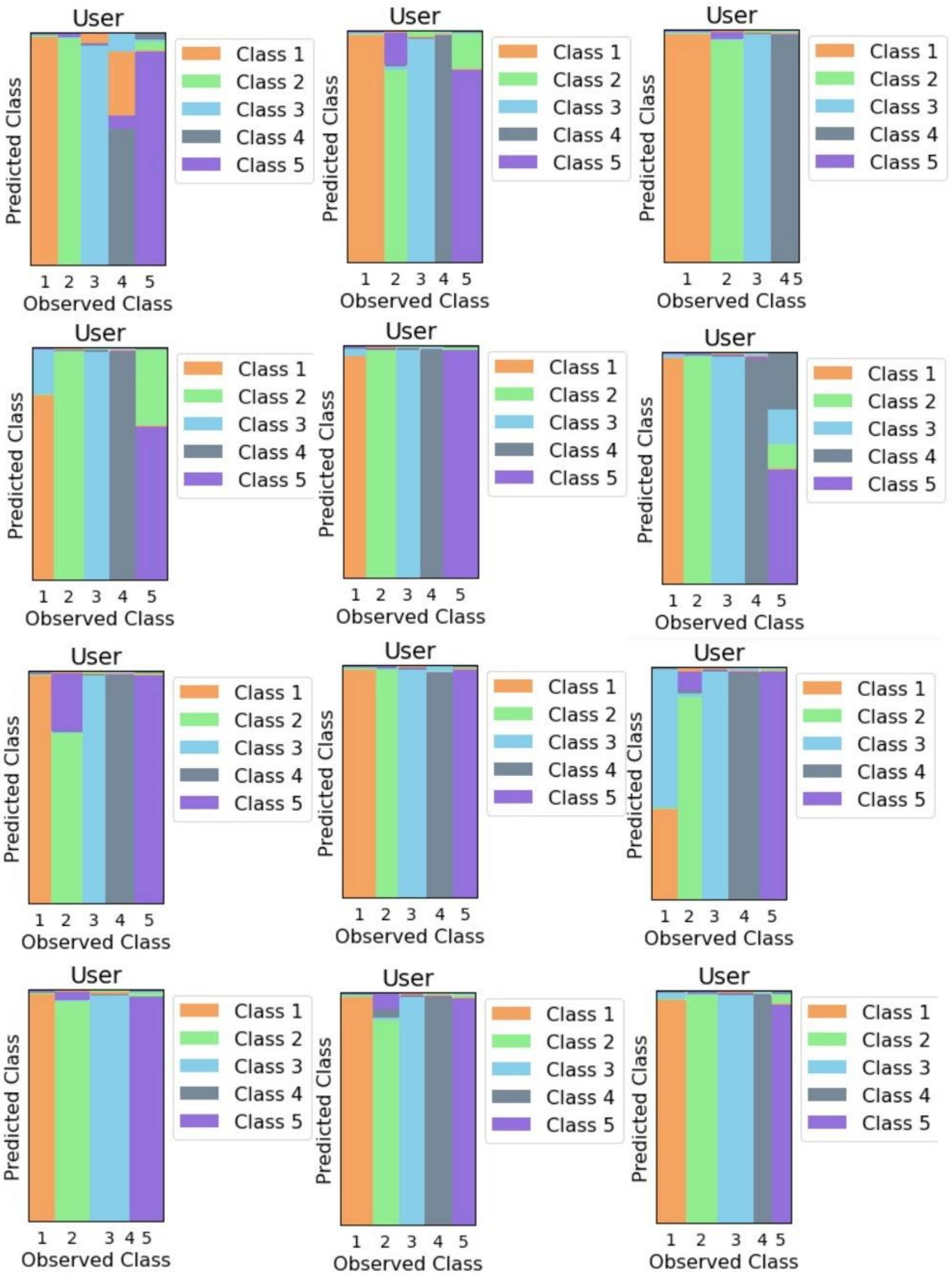
```
The F1 score for the Voting Classifier algorithm is: 0.9436
The Accuracy score for the Voting Classifier algorithm is: 0.9478
```

Results for Support Vector Machine Algorithm:



The mosaic plot of all 12 users are displayed starting from 0 in the top left to 14 at the bottom right.

Results for Voting Classifier Algorithm:



The mosaic plot of all 12 users are displayed starting from 0 in the top left to 14 at the bottom right.

Inferences:

1. The dataset had a huge percentage of missing values and outliers.

2. Treating the missing values was one of the first challenges faced. Then dropping the features, comparing the performance of the model iteratively gave us a good sense of the feature importance segment.

3. Second challenge was the outliers and their proper treatment gave the model a significant boost.

4. The confusion matrix obtained at every step for every user gave us a proper intuition of which users were performing good and which were not.

5. The user's performance was also seen to be class dependent i.e. the users performing poorly in predicting one class was seen performing well in predicting the other classes.

6. It was also seen that various models had different effects on different users.

7. Some of the users were consistently performing good irrespective of the algorithm chosen, whereas some were performing poorly.

8. It is seen that the 2nd User was performing poorly everywhere, but not in predicting all classes. The 2nd user was under performing in predicting the 5th Class for Random Forest but the same model poorly predicted 1st Class for the XGBoost.

9. On the other hand, the 10th and the 13th Users performed well in predicting all the classes irrespective of the algorithm chosen.

10. The model's performance in both the algorithms were approximately similar.

11. The reason for the poor performance by some users can be due to numerous reasons, some of them are:
    a. Probable overlap of the co-ordinates.
    b. Wrong assignment of the co-ordinates to a particular feature.
    c. Same pattern of movement while performing two different gestures.

12. We could also have tried to drop that particular class for that User, but that would have created imbalance within the User itself. Also, there would have been biasness for that data along with data loss for our model.

13. As already mentioned all the ensemble algorithms performed very similarly, but the Voting Classifier in the end gave the most robust results.

14. The performance of every algorithm used can be summarized in a table as shown below, where F1 score of each algorithm is mentioned.

	F1 Score in Percentage
Logistic Regression	66.35
Random Forest	89.61
Gradient Boost	89.05
XGBoost	89.08
Support Vector Machines	91.82
Voting Classifier	94.76

Closing reflections:

Some of the algorithms used in our project made us realize that computational capacity is something we need to consider before running these algorithms.

We did learn during this process and tried overcoming it, however considering time constraint we couldn't explore more models on this dataset.

Also, this dataset is something new to our understanding because of its highly cardinal structure, we had to look beyond traditional ways of EDA. It took thorough efforts to first understand data and then get our concepts in line with this data.

Dealing with missing value imputation was tough, as there were columns with more than 80% missing values, and since the data was cardinal, guesswork was impossible nor we could impute based on any factual data. So we had to apply all the imputation methods and see which method was giving us the best initial results.

The model can be further honed and go to prototype stage if we apply deep learning algorithms.

Considering computational limitations, we couldn't explore more many models.

Mentor: Mr. Koneti Naveen Kumar

Submitted By: Jayesh Abhishek, Gaurav Kautish, Naman Virmani, Shivam Taneja, Vivek Chauhan