MACHINFY
Story of Data

Human Activity Recognition using smart phones

Prepared by

Asmaa Shehata

Special Thanks for the most awesome and professional engineers

Engineer-**Sayed Omar**
Engineer-**Abdo Gharib**

Engineer-**Shady Nagy**
Engineer- **Doaa Mohamed**

Agenda

- Asmaa Shehata-Sr.Ex. Human Resources
- Human activity recognition using smart-phones
- Technical Explanation
  - Dataset
  - Pre-processing techniques- No Preprocessing needed for this data
  - Data Visualization
  - Model Selection and Evaluation
- Solution Impact
- Conclusion


DATA
SCIENCE

Definition of Data Science-DS

Let's understand Data science from the basics

Main Roles:

Predict

Suggest

Deploy

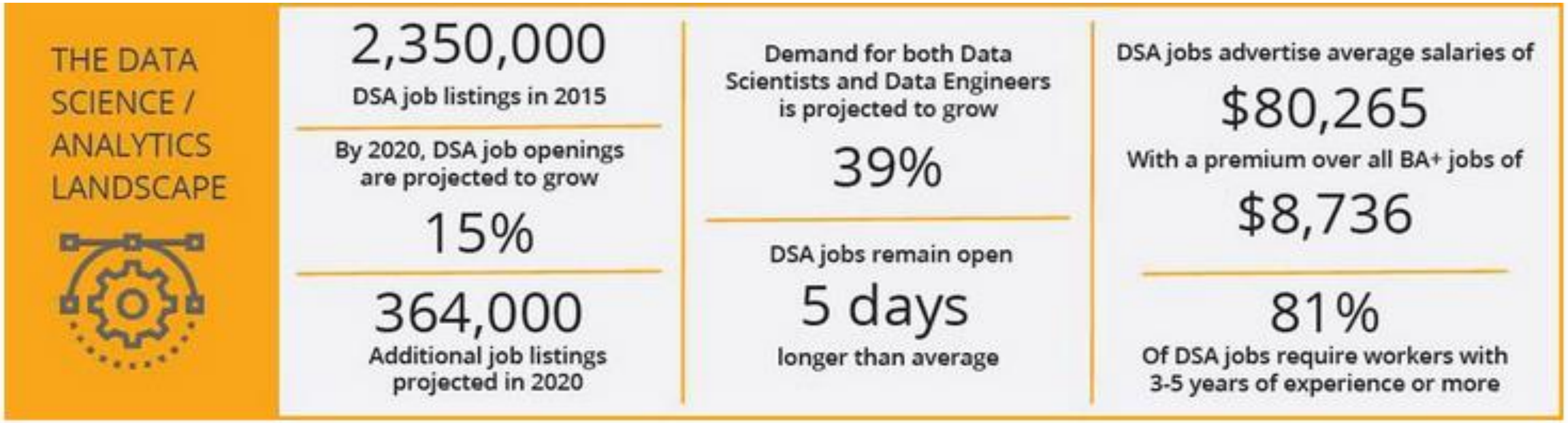
Data science seeks to reveal the hidden patterns within the data sets. Its insights are used for companies to;

- Enhance the decision-making process,
- Streamlining sales, marketing strategies and scaling the revenues
- Data science has enabled industries to create better products tailored specifically for customer experiences. for example, Recommendation Systems used by e-commerce websites provide personalized insights to users based on their historical purchases.

Article On Forbes

resource: https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#2bea8e4f7e3b

By 2020 the number of Data Science and Analytics job listings is projected to grow by nearly 364,000 listings to approximately 2,720,000




وظائف
ستندثر
من
المستقبل

Technical Explanation

- Business Case
- Objective
- Data Set
- No Preprocessing needed for this data
- EDA

Human activity recognition using smart phones- Objective

-Human activity recognition using smart-PH has wide applications in the following fields;

- health-care domain
- -Especially in elder care support, rehabilitation assistance, diabetes, and cognitive disorders.

Also on body sensing and smart PH- can measure several attributes including vital signs (e.g. heart rate, body temperature, and blood-pressure).

A huge amount of resources can be saved if sensors can help caretakers record and monitor the patients all the time and report automatically when any abnormal behavior is detected.

- security in surveillance system

Banks, public stations, hospitals…etc., so they can detect any suspicious actions and react upon it

- Smart home-integrated techniques and sensors

*It is possible to predict the activities performed by its residents based on the sensors signals along with other relevant aspects such as:-

-automatically pre-heating the coffee machine.

-controlling room lighting and temperature.

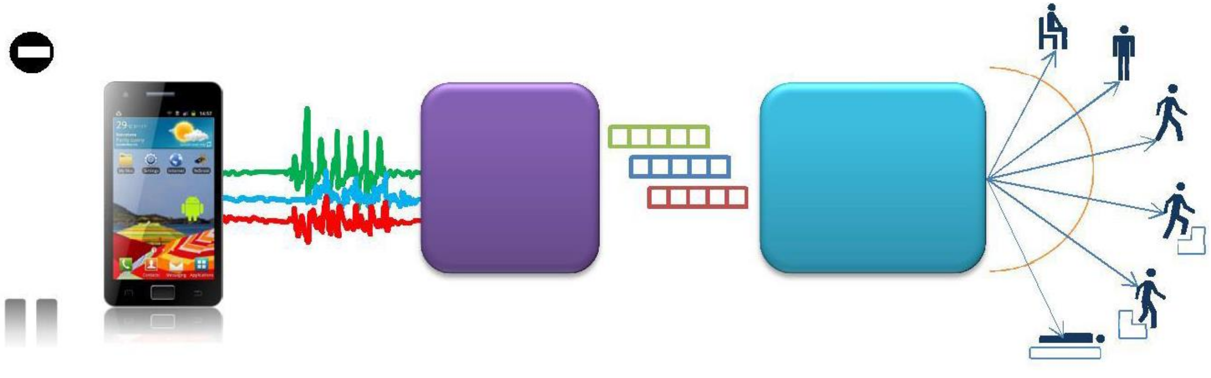
Human Activity Recognition using On-body Sensing

Human Activity Recognition-using smart phones

How HAR-Works

First elements that have to be taken into considerations are choosing sensors that will measure the human activity. Thus, in designing and recognizing the HAR- we will be using 2 major sensors embedded in any smartphone called:-

- **Accelerometer:** measures the experienced physical acceleration of an object. It has been employed for several applications in science, medicine, engineering, and industry such as for measuring vibrations in machinery, acceleration in high-speed vehicles, and moving loads on bridges.
- **Gyroscope:** is a sensing device for measuring orientation or direction. it has been introduced in electronic devices (e.g. Smartphone, game consoles) for enhancing user interfaces and gaming experience. For-HAR, this sensor has been employed in various applications such as for the detection of activities (e.g. walking, sitting , laying) and transitions between postures(e.g. whether the person is walking upstairs or downstairs)



Data Set & EDA

Train part & its shape (7352, 563)

	tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z	tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z	tBodyAcc-mad()-X	tBodyAcc-mad()-Y	tBodyAcc-mad()-Z	tBodyAcc-max()-X	tBodyAcc-max()-Y	tBodyAcc-max()-Z	tBodyAcc-min()-X	tBodyAcc-min()-Y	t
0	0.288585	-0.020294	-0.132905	-0.995279	-0.983111	-0.913526	-0.995112	-0.983185	-0.923527	-0.934724	-0.567378	-0.744413	0.852947	0.685845	
1	0.278419	-0.016411	-0.123520	-0.998245	-0.975300	-0.960322	-0.998807	-0.974914	-0.957686	-0.943068	-0.557851	-0.818409	0.849308	0.685845	
2	0.279653	-0.019467	-0.113462	-0.995380	-0.967187	-0.978944	-0.996520	-0.963668	-0.977469	-0.938692	-0.557851	-0.818409	0.843609	0.682401	
3	0.279174	-0.026201	-0.123283	-0.996091	-0.983403	-0.990675	-0.997099	-0.982750	-0.989302	-0.938692	-0.576159	-0.829711	0.843609	0.682401	
4	0.276629	-0.016570	-0.115362	-0.998139	-0.980817	-0.990482	-0.998321	-0.979672	-0.990441	-0.942469	-0.569174	-0.824705	0.849095	0.683250	

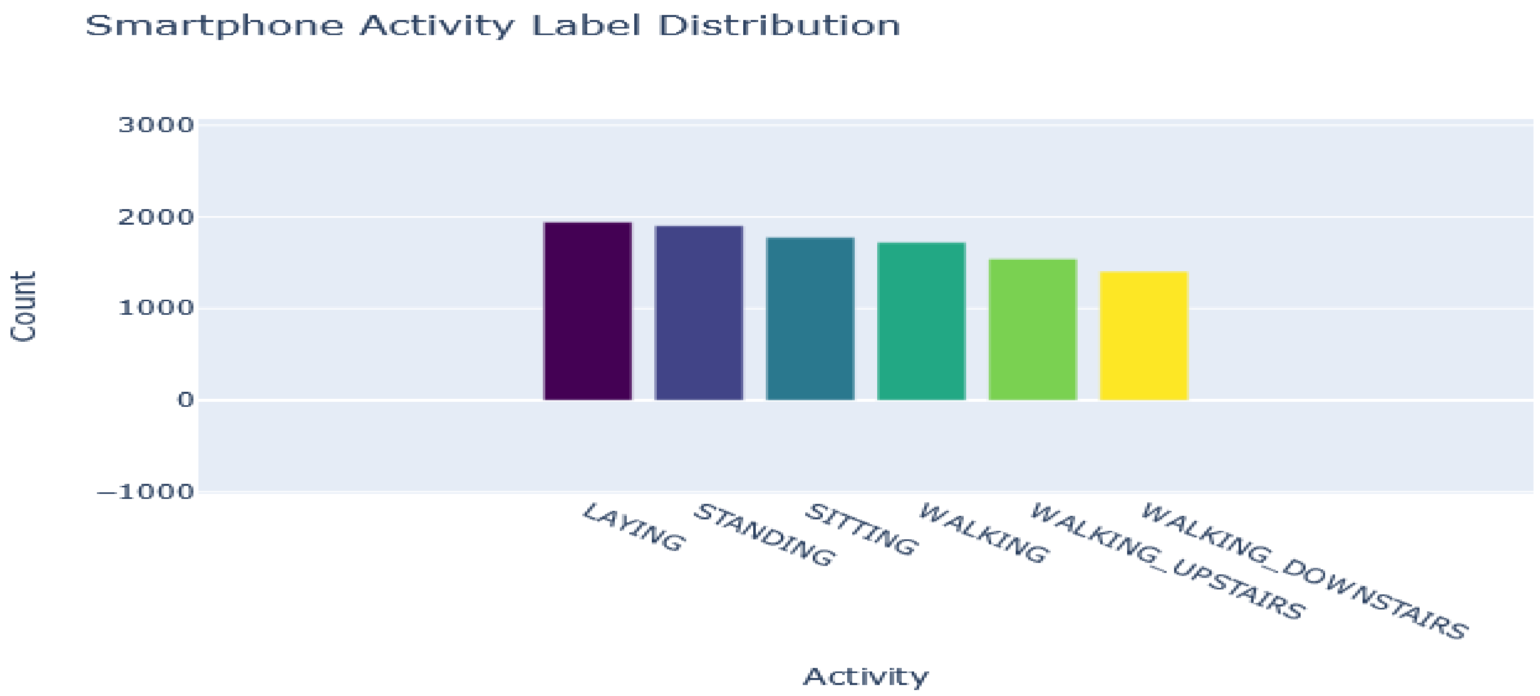
Test part & its shape(2947, 563)

	tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z	tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z	tBodyAcc-mad()-X	tBodyAcc-mad()-Y	tBodyAcc-mad()-Z	tBodyAcc-max()-X	tBodyAcc-max()-Y	tBodyAcc-max()-Z	tBodyAcc-min()-X	tBodyAcc-min()-Y
0	0.257178	-0.023285	-0.014654	-0.938404	-0.920091	-0.667683	-0.952501	-0.925249	-0.674302	-0.894088	-0.554577	-0.466223	0.717208	0.635502
1	0.286027	-0.013163	-0.119083	-0.975415	-0.967458	-0.944958	-0.986799	-0.968401	-0.945823	-0.894088	-0.554577	-0.806013	0.768031	0.683698
2	0.275485	-0.026050	-0.118152	-0.993819	-0.969926	-0.962748	-0.994403	-0.970735	-0.963483	-0.939260	-0.568512	-0.799116	0.848305	0.667864
3	0.270298	-0.032614	-0.117520	-0.994743	-0.973268	-0.967091	-0.995274	-0.974471	-0.968897	-0.938610	-0.568512	-0.799116	0.848305	0.667864
4	0.274833	-0.027848	-0.129527	-0.993852	-0.967445	-0.978295	-0.994111	-0.965953	-0.977346	-0.938610	-0.560831	-0.825894	0.849179	0.670700

EDA or- Data Visualization

- **Balance of Activity Classes**

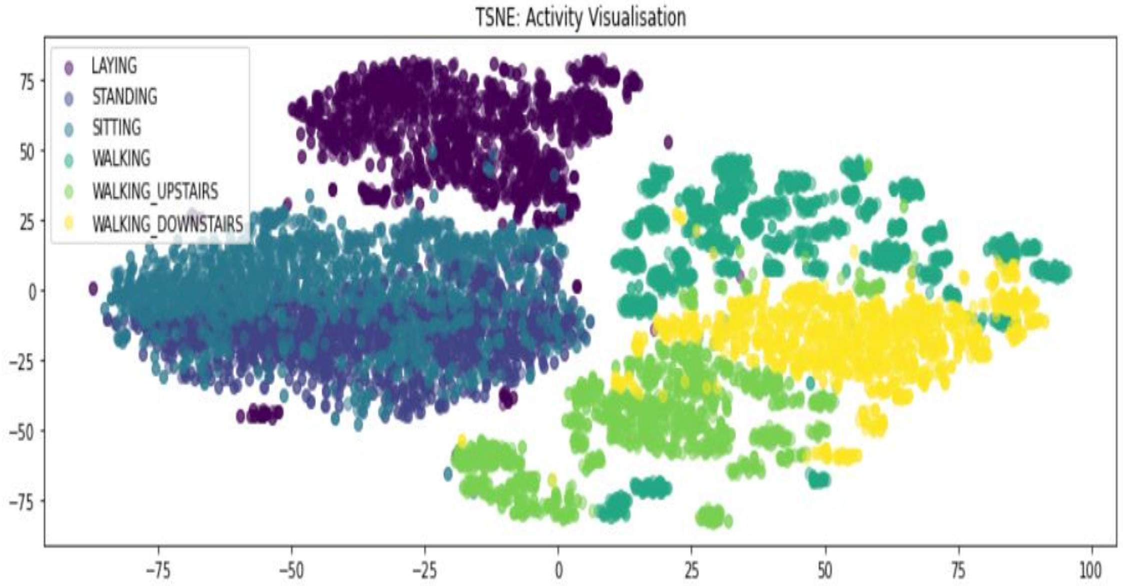
- According to the data set of HAR- we in the first place checked how balanced the data sets are with respect to the different activity labels on the 30 subjects. We find from the figure below that the volumes are well balanced which making it easier to model, and confirm that we are correctly loading and interpreting the dataset.

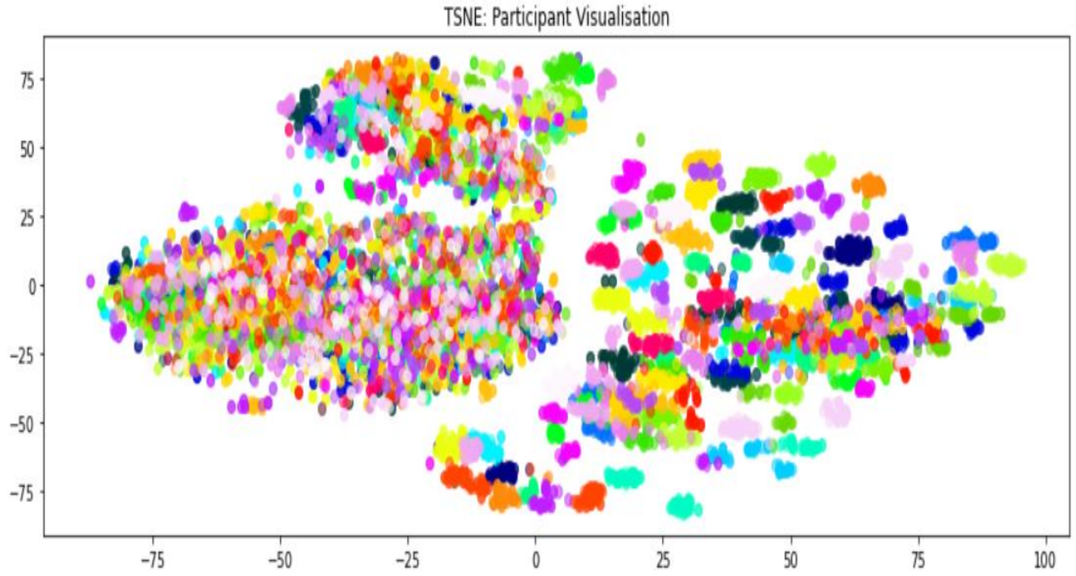


EDA- Cont.

To reduce data complexity, time consuming and organize the data, we can use the T-SNE technique for dimensionality reduction with the multivariate dataset .

personal information of the participants. Everybody has for example an unique/sparable walking style (on the upper right). Therefore the smartphone should be able to detect what you are doing and also who is using the smartphone (if you are moving around with it).

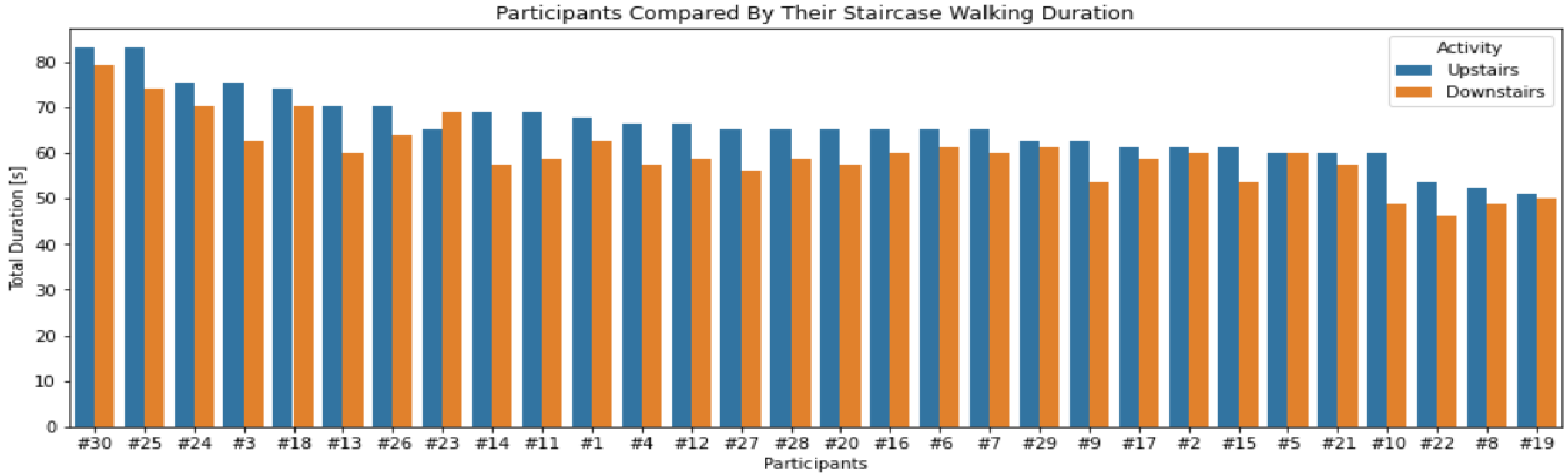




T-SNE 's been used to separate the activity. It will be also used to separate each and every participant so we can determine each and every activity for each participant through their movements which will be read from the accelerometer and gyroscope sensors

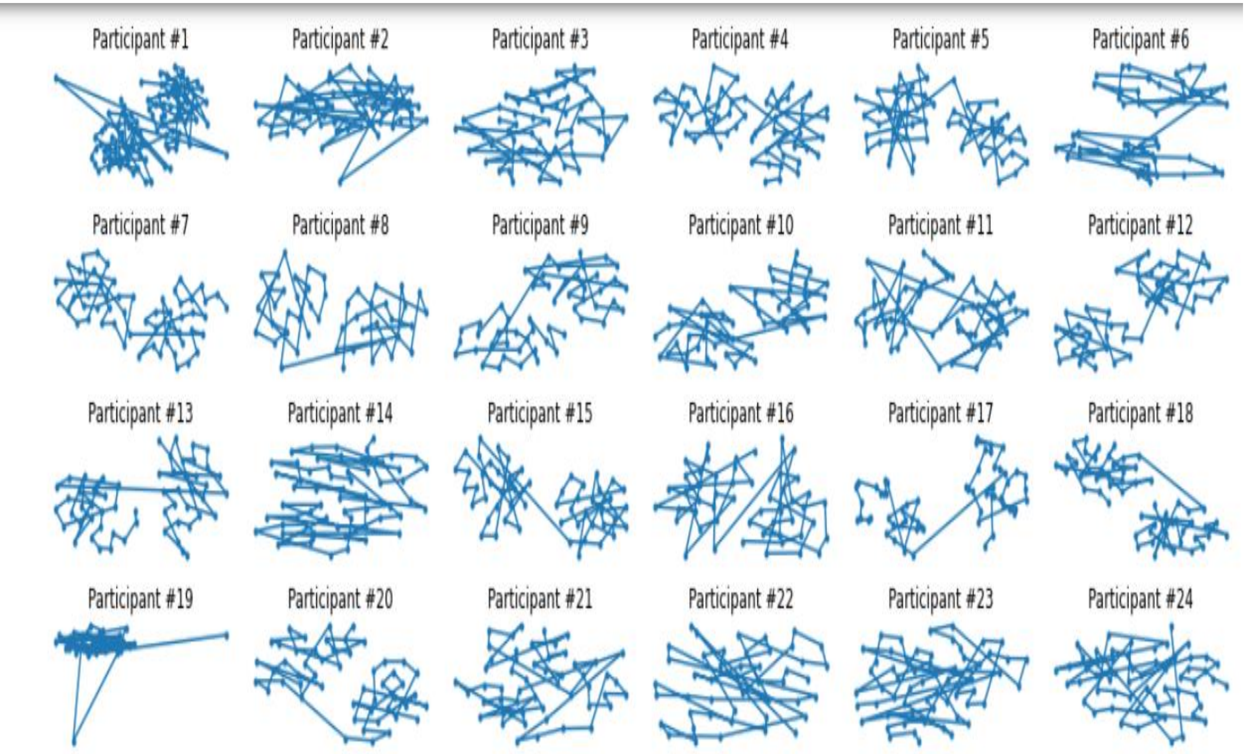
EDA- Cont.

all participants have more data (consuming more time) for walking upstairs than downstairs. Assuming an equal number of up- and down-walks the participants need longer walking upstairs.

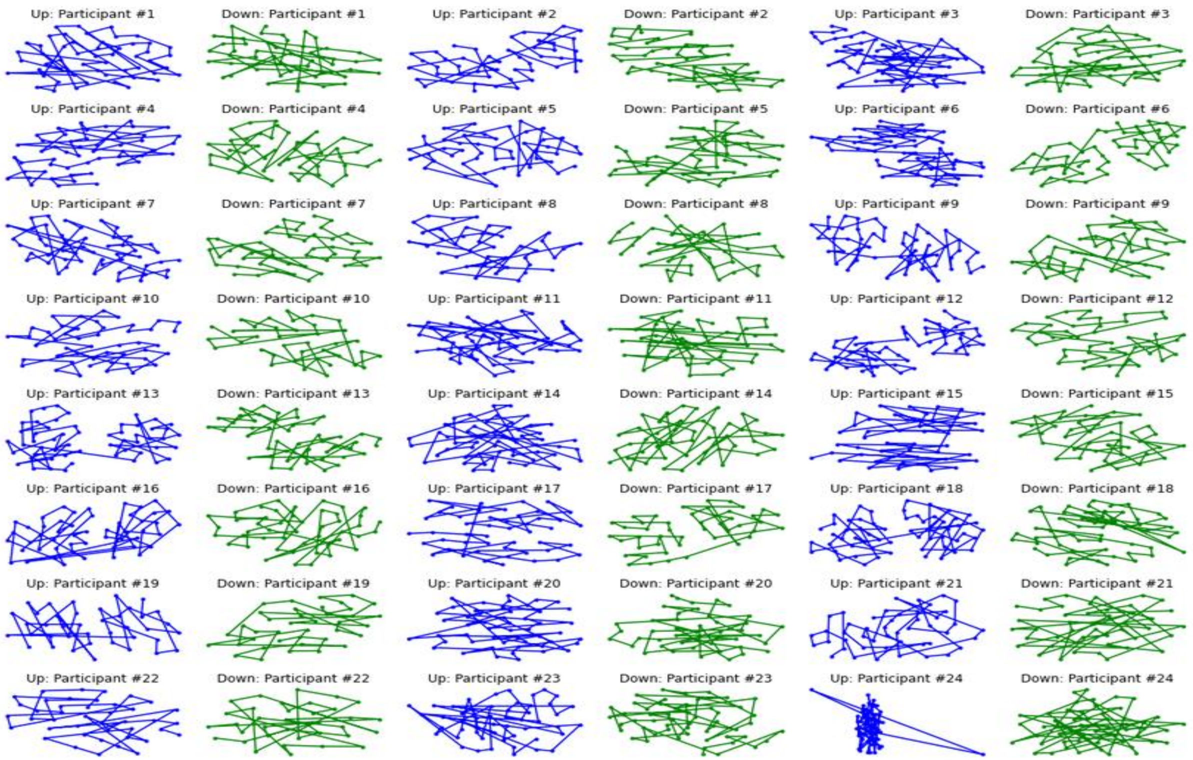


EDA- Cont.

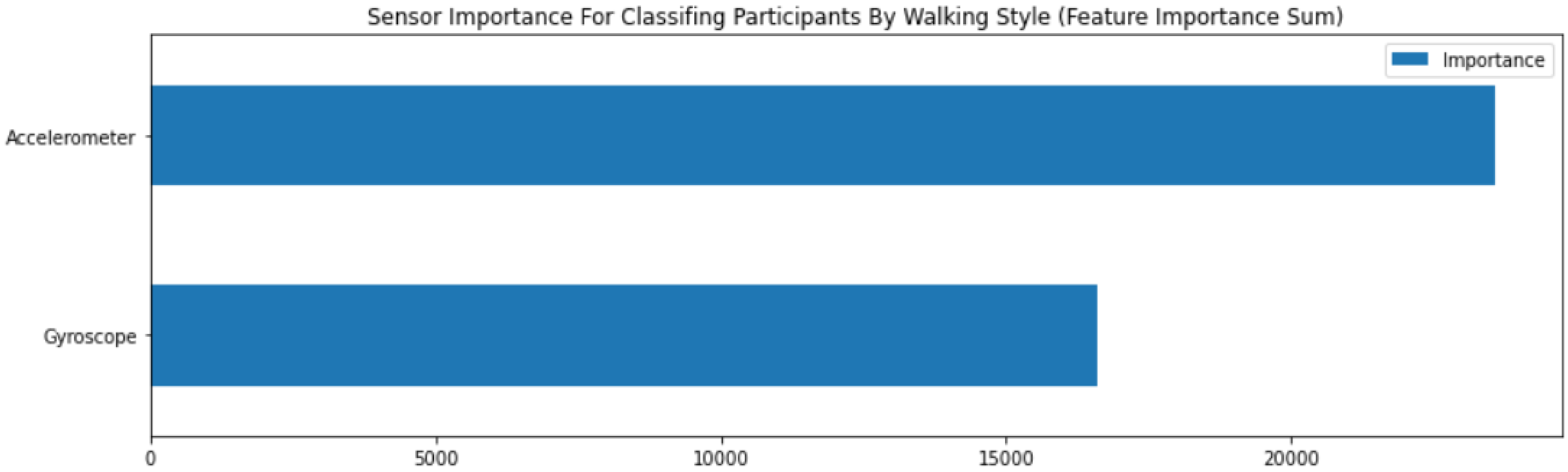
- Visualizing the walking structure for each participant.



Staircase Walking Style For Each Participant(Up & Down)



EDA- Cont.- Feature Importance



Model Building

The process of designing an effective model building:-

1-Define the problem or the business case

2- Gather the data

3-Split the data correctly

4- Develop a benchmark algorithm

5- Create the confusion matrix & F1- score (Model Evaluation)

6- Activity prediction-integrated system (Deployment)

Building Model Cont.

- **2- Gather the data**

	tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z	tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z	tBodyAcc-mad()-X	tBodyAcc-mad()-Y	tBodyAcc-mad()-Z	tBodyAcc-max()-X	tBodyAcc-max()-Y	tBodyAcc-max()-Z	tBodyAcc-min()-X	tBodyAcc-min()-Y
0	0.288585	-0.020294	-0.132905	-0.995279	-0.983111	-0.913526	-0.995112	-0.983185	-0.923527	-0.934724	-0.567378	-0.744413	0.852947	0.685845
1	0.278419	-0.016411	-0.123520	-0.998245	-0.975300	-0.960322	-0.998807	-0.974914	-0.957686	-0.943068	-0.557851	-0.818409	0.849308	0.685845
2	0.279653	-0.019467	-0.113462	-0.995380	-0.967187	-0.978944	-0.996520	-0.963668	-0.977469	-0.938692	-0.557851	-0.818409	0.843609	0.682401
3	0.279174	-0.026201	-0.123283	-0.996091	-0.983403	-0.990675	-0.997099	-0.982750	-0.989302	-0.938692	-0.576159	-0.829711	0.843609	0.682401
4	0.276629	-0.016570	-0.115362	-0.998139	-0.980817	-0.990482	-0.998321	-0.979672	-0.990441	-0.942469	-0.569174	-0.824705	0.849095	0.683250

Data Shape
(10299, 563)

Building Model Cont.

3-Split the data

Note: This data-set was luckily split into 2 parts;

*Train part & its shape (7352, 563)

	tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z	tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z	tBodyAcc-mad()-X	tBodyAcc-mad()-Y	tBodyAcc-mad()-Z	tBodyAcc-max()-X	tBodyAcc-max()-Y	tBodyAcc-max()-Z	tBodyAcc-min()-X	tBodyAcc-min()-Y	t
0	0.288585	-0.020294	-0.132905	-0.995279	-0.983111	-0.913526	-0.995112	-0.983185	-0.923527	-0.934724	-0.567378	-0.744413	0.852947	0.685845	
1	0.278419	-0.016411	-0.123520	-0.998245	-0.975300	-0.960322	-0.998807	-0.974914	-0.957686	-0.943068	-0.557851	-0.818409	0.849308	0.685845	
2	0.279653	-0.019467	-0.113462	-0.995380	-0.967187	-0.978944	-0.996520	-0.963668	-0.977469	-0.938692	-0.557851	-0.818409	0.843609	0.682401	
3	0.279174	-0.026201	-0.123283	-0.996091	-0.983403	-0.990675	-0.997099	-0.982750	-0.989302	-0.938692	-0.576159	-0.829711	0.843609	0.682401	
4	0.276629	-0.016570	-0.115362	-0.998139	-0.980817	-0.990482	-0.998321	-0.979672	-0.990441	-0.942469	-0.569174	-0.824705	0.849095	0.683250	

*Test part & its shape(2947, 563)

	tBodyAcc-mean()-X	tBodyAcc-mean()-Y	tBodyAcc-mean()-Z	tBodyAcc-std()-X	tBodyAcc-std()-Y	tBodyAcc-std()-Z	tBodyAcc-mad()-X	tBodyAcc-mad()-Y	tBodyAcc-mad()-Z	tBodyAcc-max()-X	tBodyAcc-max()-Y	tBodyAcc-max()-Z	tBodyAcc-min()-X	tBodyAcc-min()-Y
0	0.257178	-0.023285	-0.014654	-0.938404	-0.920091	-0.667683	-0.952501	-0.925249	-0.674302	-0.894088	-0.554577	-0.466223	0.717208	0.635502
1	0.286027	-0.013163	-0.119083	-0.975415	-0.967458	-0.944958	-0.986799	-0.968401	-0.945823	-0.894088	-0.554577	-0.806013	0.768031	0.683698
2	0.275485	-0.026050	-0.118152	-0.993819	-0.969926	-0.962748	-0.994403	-0.970735	-0.963483	-0.939260	-0.568512	-0.799116	0.848305	0.667864
3	0.270298	-0.032614	-0.117520	-0.994743	-0.973268	-0.967091	-0.995274	-0.974471	-0.968897	-0.938610	-0.568512	-0.799116	0.848305	0.667864
4	0.274833	-0.027848	-0.129527	-0.993852	-0.967445	-0.978295	-0.994111	-0.965953	-0.977346	-0.938610	-0.560831	-0.825894	0.849179	0.670700

Building Model Cont.

- **4- Develop a benchmark algorithm**

For this data-set, 6 different algorithms have been chosen.
accordingly we can select the most effective model with the highest accuracy

- 1- LGBM CLS
- 2- LR CLS
- 3- SVM CLS
- 4- KNN CLS
- 5-Decision tree CLS
- 6- Random Forest CLS

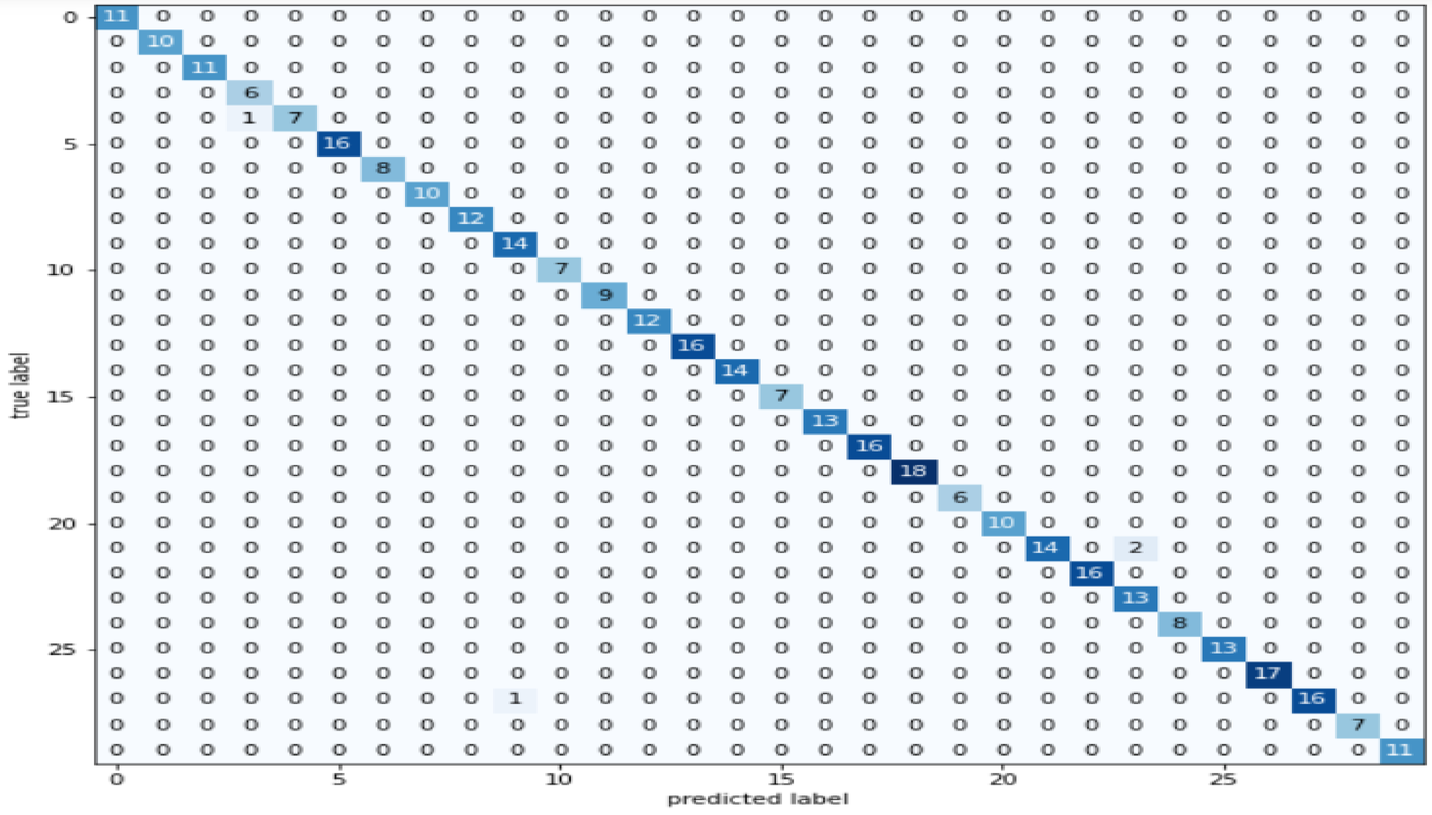
Building Model Cont.

-Logistic Regression(LR) has been chosen because;

- Accuracy on Activity WALKING: 50%
- Accuracy on Activity WALKING_UPSTAIRS: 99%
- Accuracy on Activity WALKING DOWNSTAIRS: 99%
- Accuracy on Activity STANDING: 72%
- Accuracy on Activity SITTING: 61%
- Accuracy on Activity LAYING: 64%

5-Model Evaluation.

- Confusion Matrix



LR F1-score: 98%
LR Recall: 98%

6-Activity prediction-integrated system

Let’s see How it works