

FACULTY OF SCIENCE, ENGINEERING AND COMPUTING

**MSc DEGREE
IN
MSc Information systems**

Name: Padmavathi Palakala

ID Number:2006093

**Project Title: Project Initiation and Implementation of Human Activity Recognition
System**

Date:05/01/2021

**Supervisor1: Dr Islam Choudhury
Supervisor2: Al-Kfairy, Mousa A**

Kingston University London

WARRANTY STATEMENT

This is a student project. Therefore, neither the student nor Kingston University makes any warranty, express or implied, as to the accuracy of the data or conclusion of the work performed in the project and will not be held responsible for any consequences arising out of any inaccuracies or omissions therein.

Table of Contents

Abstract	8
Acknowledgement	10
Chapter 1	11
Introduction	11
1.1 Background & Motivation.....	12
1.2 Research problem.....	12
1.3 AIM.....	13
1.4 Objectives	13
1.5 Research Questions	13
1.6 Ethical	14
Chapter 2	15
Literature Review	15
2.1 Introduction	15
2.3 Advantages and disadvantages of machine learning	16
2.4 Advantages.....	17
2.5 Disadvantages.....	17
2.6 From the Human Activity Recognition -Point of View	18
2.7 Scope.....	18
2.8 Limitations	19
2.9 Historical review of events in Human Recognition through machine learning:	19
2.10 Conclusion:	27
2.11 Literature review Futuristic scope to suite the business.....	29
Chapter 3	31
Business Modelling for Project	31
3.1 Mission statement	31
3.2 Identification of IT project	31
3.3 Producing requirements Catalogue	33
Chapter 4	36
Project Initiation.....	36

Prince2 approach producing a PID document with a convincing project plan	36
4.1 Method description.....	38
4.2 Requested total budget.....	39
4.3 Cost benefit analysis.....	40
4.4 Project breakdown structure	40
4.4 Brief description of the entire process.....	42
4.5 Gantt chart and critical path analysis	42
Chapter 5	45
Methodology and Design Human Activity Recognition System	45
5.1 Hardware used	46
5.2 Understanding of DATA.....	47
5.3 DATA Preparation	48
5.4 Cleaning of the DATA.....	48
5.5 Transformation of DATA.....	50
True Positive, True Negative, False Positive and False Negative.....	52
Chapter 6	53
Implementation.....	53
6.1 Load DATASET	53
6.2 Splitting of the data	53
6.3 Build the model	53
6.4 Exploratory data Analysis (EDA).....	54
Chapter 7	72
Conclusion	72
7.1 Results.....	72
7.2 Research gaps.....	74
References.....	75
Liberty, E., Lang, K., & Shmakov, K. (2016). Stratified sampling meets machine learning. In International conference on machine learning (pp. 2320–2329).....	77

List of Tables

Table 1 Project plan	38
Table 2 Requested project budget.....	39
Table 3 Gantt Chart.....	43
Table 4Table 4 x_train dataframe	55
Table 5Table 5: y_train Dataframe	56
Table 6 Logistic Regression Classifier	63
Table 7Naive Bayes Classifier.....	64
Table 8 Random Forest Classifier.....	67
Table 9 K-Fold Cross Validation Classifier.....	68

List of Figures

Figure 1 Scope diagram of HAR System.....	32
Figure 2 Use case diagram of HAR System	34
Figure 3 Project initiation plan document or PID using Prince2 approach	37
Figure 4 project breakdown structure	41
Figure 5 Critical path analysis of the project	44
Figure 6 Different phases in CRISP-DM process Model	46
Figure 7 Confusion Matrix.....	52
Figure 8 Heat Map	57
Figure 9 Bar Graph	58
Figure 10 Line Graph.....	58
Figure 11DataFrames.....	59
Figure 12 Line Graph.....	60
Figure 13 Histogram of features	61
Figure 14 All list of Features	61
Figure 15 KNN Classifier	65
Figure 16 SVM Classifier	66
Figure 17Scatter Plot.....	68
Figure 18 Keras Model	71
Figure 19 Accuracy of Classifiers.....	73

Abbreviations

HAR- Human Activity Recognition

UML- Unified Modelling Language

USP -Unique selling point

AbHAR - Abnormal Human Activity Recognition

ARS - Activity Recognition System

RGB - Red Blue Green Camera

SWOT - Strengths Weakness Opportunities and Threats

PESTLE - Political Economic Social Technological Legal Environmental

CRISP-DM - Cross Industry Standard Process for Data Mining

PID - Project Initiation Document

CBA - Cost Benefit Analysis Technique

RAM - Random Access Memory

SVM - Support Vector Machine

N/A - Not Applicable

MNAR - Missing not at Random

MAR - Missing at Random

2D – 2 Dimensional

3D - 3 Dimensional

PCA - Principal Component Analysis

EDA - Exploratory Data Analysis

KNN – K-Nearest Neighbor

**PROJECT INITIATION AND IMPLEMENTATION OF HUMAN ACTIVITY RECOGNITION
SYSTEM**

Abstract

The Modern man is becoming busy and lazy day by day which may not look like a good sign; however this is giving more and more scope of innovation and improvisation where the man is becoming smarter and developing smarter machines or developing smarter techniques, which has resulted in the Artificial intelligence where machines are thinking and acting smart and replacing human efforts and making human life much easier. Human Activity recognition (HAR) is one of the outcomes of this advance technology, HAR enables a wide range of pervasive computing applications by recognizing the activities performed by the human (user), with HAR predicting the accurate activity is utmost importance, this accuracy depends on the algorithm that is used if the inappropriate algorithm is used it may hinder the accuracy results of HAR system.

In the first chapter of this paper the aims and objectives involves the research of various architecture and design processes of the human activity detection process which involves machine learning and python coding. Non-functional and functional requirement highlights the basic requirement of the overall system which helps to make an error-free and risk-free system

In the present context of the, Alpha Bread company which is a bread making factory which is well known for their hand made Garlic breads this company is growing in size day by day, however the management of the company is facing an issue with the employees behaviour and their skill of bread making, employees are mostly not at their workstation during the production hours and it is becoming very difficult for the line managers to monitor them, hence one of the executive leaders proposed to have a Human Activity Recognition system implemented in the factory which will monitor the presence of the employee at the workstation and also monitors their hand movement while making the bread which will affect the outcome of the bread quality and the projects main aim is to find out the idle time of an employee where he is not been productive this can be calculated calculating the time between any two activities of an employee. So, the management has agreed to initiate a project on implementing a HAR system in the factory. As the process of pre project initiation the IT team would be working on the various UML diagrams to build a system and the operations managers would follow prince2 model to initiate the project

The human activity recognition project of the organization Alpha Bread follows the technical application of Python and machine learning programming concept. linear regression techniques

are some important parts of the project work that help to identify the staff idle time of this company. Development of the decision tree model in the machine learning software by using the staff working time dataset of the bread making company helps to meet the research aim in the data management section of this analysis.

In the third chapter the overall analysis of the key aspects of the specific project has been highlighted. The Agile project development part focuses on the project working structure and plan analysis phase. This part develops the idea about the Prince2 method to implement the project initial planning document. The project planning section and project budget analysis helps to integrate the project development quality. Cost benefit analysis and budget forecasting with project scheduling factor helps to meet the successful project completion goals. Project breakdown structure and risk analysis are some effective parts of this section.

Acknowledgement

I would like to take this opportunity to thank all those people who have extended their support and provided their advice and encouraged me. I would like to acknowledge with gratitude, and convey my gratitude to my professor, for his guidance and assistance whenever I needed any help during the full course of my study. I would also thank all those authors who has made their study available for me to read and research upon. Last but not the least my family and peers for their massive support during the conduction of the entire research study.

Thanks, and regards,

Chapter 1

Introduction

In today's world internet is spreading like air across the globe and it has become oxygen to businesses, no business can endure without internet. In addition to this software development is helping the business transition from many traditional methods to advance methods which can increase the productivity, built a better product, improve the quality and manage human resources. Over a period of time there are many tools, techniques and methodologies developed to help the business to improve their operations. Technology has taken a big leap after the introduction of Artificial Intelligence (AI) where the technology has transformed itself drastically and now it is able to think by itself almost like a human being, this artificial intelligence is developed with the help of many algorithms such as machine learning and deep learning these techniques will help the computer to analyse the unstructured raw data, images and videos and transform it into a meaningful information which can be used in decision making, HAR is used in various fields such as Sports, medical, elder care, automated surveillance military, fitness tracking, military, etc... where it is doing wonders.

In the present context of this paper, we will understand and utilize various management techniques and software developments which will help the Alpha breads company thrive and also address one of its concern that is improving the efficiency of employees by monitoring their activities with the help of human activity recognition using machine learning

The system has been developed with an objective of assessment of positions of the employees in the production house. The system has included machine learning and the software system which is effective in terms of development and assessment of the employee performance. This can be linked to the quality of work in the company and also the relevant assessments of the company are possible based on employee presence in the production. The work environment of the company is also required to be developed based on system assessments. The system has been derived from the background situation in the company and the objectives are also developed based on the research.

1.1 Background & Motivation

Alpha Breads is an emerging bread making company which has a production unit (warehouse) in England which has grown in size in the past few years, they have earned reputation for their unique handmade Garlic breads, which bought them a unique identity in the market. Alpha Breads are supplied to all reputed supermarket across London and now they are trying to stretch their reach to other parts of England. currently issues with their employees' performance as the employees are become irresponsible and not doing their basic duties in the absence of supervisors and some of the employees are not delivering a quality product because of the poor skills of handmade bread making these two issues are affecting the quality of the finished good. As the USP (Unique selling point) of Alpha breads is their hand made garlic breads the management is a bit concerned. The bread making employees are not performing up to the mark due to which there is a decrease in production and decline in the quality of the product as the Alpha bread is known for their hand made Garlic breads due to the quality issue their brand image is getting effected. Due to which the management wants to monitor employee's performance with the help of human activity recognition and help them improve their bread making skills and improve their efficiency

1.2 Research problem

There are many challenges and issues in the selection of the appropriate techniques and tools for data gathering, manipulation and storage, there might be a problem in selecting the right algorithm in performing the predictions as it is very important to capture the inter class similarity and variability. We might face a problem in availability of time and storing the data and resource constraints as processing power. We might also need to deal with the system accuracy, latency and processing power. We might see a lot of issues during the initial stages of HAR as in picking the right sensor, choosing the metrics and attributes to be measured, need to place the sensor at the appropriate location. Above all there might be problem with the user acceptance of this experiment as it should not hinder his activities and rather help the user so that the user can voluntarily extend his involvement

1.3 AIM

The research aim is to initiate project of implementing Human Activity Recognition system for effective monitoring system to determining the presence and absence employees on the production floor and observe their activities during the production hours while the managers are out with the help of wearable sensors and machine learning processes.

1.4 Objectives

- To develop a machine learning system to detect the presence and absence of the employee in the production section
- To determine deep learning for developing a system with sensors to detect the motion of the employees
- To determine the accuracy of the HAR using different machine learning and deep learning algorithms
- To accumulate system that is effective in the context of human activity recognition and can sense a specific person from its digital identity during a particular time frame
- To determine the system with python coding and explain these with several diagrams and tables and codes

1.5 Research Questions

The questions that this research addresses are listed below.

- 1) How to develop a machine learning system that can be able to differentiate the presence or absence of the human being within the production system?
- 2) How to detect the physical motion of the employees within a workplace by only using motion detecting sensors?
- 3) How can a system be accumulated in order to recognize activities of the human beings and how a person can be detected from the way he is active digitally?
- 4) How to explain the result of these questions by using a python coding system and codes and diagrams?

1.6 Ethical

The research is really important in order to implement an advanced level of safety within different workplaces. The implementation of sensor-based motion detectors are required to increase safety and security systems (Alshammari, 2019, p. 322). Security is one of the most important and key factors in every place. From business to industry, it is highly required to maintain a real safe and secure environment so that employees or even people from management with bad intentions, thinks twice before performing anything that is going to be harmful for the company but also can harm their career as well as personal lives. It is impossible to understand the motives or mindset of all the company employees as different people come from different places along with different mentality (Baum, 2017, p. 762). Theft or stealing company goods has often been observed by the employees which affects the company in a bad manner. Reproductive units of a business company or industry have been noticed gone missing a huge number of times. Employees generally never give a name that might have done it or probably they just do not know about the fact. Allowing a tight security system within the company infrastructure prohibits the employees to perform this type of false work practices (Dawes *et al.* 2017, p. 990).

Implementation of tighter security systems along with motion detectors can help the security system of a company to get more precise in order to understand and determine the people who potentially can be a threat to the company along with their false practices within the workplace. There has been an idea which for several times is running within the heads of the company management that is if motion sensors are able to recognize a person from the body movement that he makes or the activities that he does (Dawes *et al.* 2020, p. 381). The mechanism on a primitive level, has been observed to work like the sensors initially records the digital movement of an individual and then creates a folder of that particular person within the system, along with time the system keeps recording and storing dates of a particular person manually till a time when the system starts to recognize the person from the action that he makes or the way he behaves. These motion sensor-controllers are to be controlled manually for some days in order to help them recording the data of the people and after a few days, after collecting about 47% of the data about any individual, the system has been seen to start recognizing people from their digital behaviour (Hummer *et al.* 2018, p. 558). Implementing sensor detectors like these not only will help to improve building a tighter security but also will analyse and identify any employee whenever and wherever required.

Chapter 2

Literature Review

2.1 Introduction

The intend of the literature review is for understanding the evolution of human activity recognition from the past to understand and utilize the expertise to use the subject knowledge to develop the project scope and see what solutions can be implemented in the business. “Alpha Bread factory” is a wholesale business of manufacturing breads. Well established and well known for their hand made breads, however the management and organization is looking forward for optimum productivity and quality from the business. Business area of interest is increasing the productivity by 120% with the help of latest technology’s and trends in which human activity recognition is one key scope of the increase in productivity. There are two things to look forward

How well we can use the human activity recognition to ensure employees are performing duty as per the policy’s and roles

With the use of Human activity recognition to increase the efficiency and productivity

Before we deep dive into the literature review of human activity recognition through machine learning lets understand what human activities are and how can we recognize them with the help of current technology and trends

While we study the literature review of human activity recognition lets go through and introduce ourselves to understand what are the human activities recognition and what is the evolution of the machine learning in this area. Through machine learning recognition of human activities plays an important role in the current dynamic era of the evolving technology which clearly defines the role of human-to-human interaction and social relations. As these actions of the human provide the identity, personality and psychological mode of human activities which is really difficult to procure or extract. The ability of recognizing the activities of other person is the core subject of study in the area of machine learning. As an input to this activity there are various research applications involved to drive i.e., video surveillance, human interaction with the computer, wearable sensors devices, android mobile phones and images.

To understand the bifurcation of activities we have two important questions to understand i.e., what action it is and where in the video. When we try to recognize the human activity, we have to regulate the kinetic state of an individual so that the system can effectively identify the activities like walking, running anything relatively easy to recognize. (Vrigkas, et al., 2015)

2.2 Context for Literature review:

Within different companies or organizations, the security cameras or the security systems are not always able to capture the face of particular employees and that is why in order to track which employee is doing what within the workplace, we have human face recognition implemented within the security systems with the help of machine learning and deep learning, it is important for several reasons. Many a time theft or stealing of company goods, as well as organizational products, takes place affecting the productivities of a particular company. Along with improved security systems, it is going to be hard for any employee with false intention to pursue his work (Siirtola, 2019, p. 5151). After implementing human activity recognition, security cameras would be able to recognize people faces. In order to do that people or employees of a company are required to record their photo image by using an embedded accelerometer and gyroscope. Linear acceleration as well as the angular velocity of the employees are required to be recorded so that recognizing them becomes easier for the security system. Except for controlling theft, this kind of security system helps to maintain a safe and secure workplace atmosphere. On a similar line with regards to our project we would be recognizing the human activities such as Walking downstairs, standing, sitting, laying, walking upstairs and walking.

2.3 Advantages and disadvantages of machine learning

Along with time, one of the best innovations that technology has to provide to mankind is machine learning. Not only it helps to bring ease to the job but also helps save a lot of time as well as human efforts. Recognizing people from their body language or behavioural aspects is one of the key aspects that machine learning has to offer and along with plenty of advantages it also carries some disadvantages along with itself which is going to be discussed here.

2.4 Advantages

The advantages are plenty. Machine learning saves manpower as well as time also the efficiency that machine learning provides is way more than the human capabilities. In case of recognizing human through their behaviours a human eye can be deceived but deceiving a machine is not going to be that easy. Machine learning in the industry can impact on analyzing and processing the prediction of the future of a product depending on the current market value of that. A huge analysis of power is what machine learning has to offer (Wuest *et al.* 2016, p. 40). It has no limits as it can keep working for hours or even days. Uninterrupted service that it provides helps a company or organization to maintain the consistency of the services that it provides. In the case of the data entry, it makes the job simpler by following a particular pattern or programmed to execute the task. With easy spam detection and more accuracy in providing services, machine learning has been hugely helping mankind.

2.5 Disadvantages

Along with plenty of advantages that machine learning has to offer, it also comes along with some of the disadvantages that almost every machine has to offer. Machine learning systems require a massive amount of data sets to be trained on and these data sets have to be in excellent quality to improve the ability of the machine to perform its best. The data are required to be unbiased as well. Even after providing enough sets of data, there is a high chance of an error occurring which is the reason, not machine learning systems are not getting accepted within every company. Interpretations of the results that are why can always not be trusted. Way too much time, as well as effort, is to be implemented in order to set a machine learning system (Mohamed, 2017, p. 7). The implemented human recognition system within security cameras is required to be tested enough so that it can be determined whether the system is providing the right result in order to identify the employees of the companies in several different systems or not.

2.6 From the Human Activity Recognition -Point of View

The usage of Human activity recognition is to improve the efficiency of the employees this can be increased by using or implementing human activity recognition within the employees of a company. Many times, it has been noticed that employees, instead of working and producing products, would just sit relentlessly being lethargic to their work. Implementation of human recognition within the security system will identify the people or employees being careless about their job increasing the productivity level of a company (Matute, 2017, p. 56). With the help of machine learning systems, several applications or programmers can be developed that can help employees in order to finish their jobs or responsibilities in a soon and more effective manner.

2.7 Scope

Along with time people are understanding importance as well as capabilities of the machine learning systems and scope in this field is gradually increasing every day. From developing more software applications in order to make work-life simpler to making different applications or programs can be done through machine learning systems which are going to be impactful in the future providing more opportunities to the people in future (Mathai, 2020, p. 3585).

As per our business point of view we have different perspective to drive the business with the help of human activity recognition. In this growing technology and scope of business we see various reasons for the usage of human activity recognition through advance usage of machine learning. To develop the business plan with less people to monitor the business rather technology doing similar activity for us. Various reasons to adopt the human activity recognition for the business are mentioned below for manufacturing unit like Alpha Bread company

- Employee login and logout to the devices
- Employees in the production unit production time gets measured
- Employees Leisure activities get monitored to drive them
- All actions getting recorded to provide feedback
- On time recognition of the human activity by machine and alarm and triggering system
- Low production units getting monitored at employee level

- Based on the reports generated appropriate appraisal mechanism need to be applied to the employees

The focus is to review if the above required actions can be really monitored through human activity recognition with machine learning. my review of literature or point to review is to know if all the mentioned activities are feasible to execute. My literature review will broadly focus on the technological developments and what all human actions can be monitored in the current time period, what all can be future scope for my development of the business

2.8 Limitations

There are still limitations of machine learning like it uses big data in order to understand a procedure and has an increased chance of error in the result depending on the job that it is performing (Kenge, 2020, p. 80).

2.9 Historical review of events in Human Recognition through machine learning:

The above-mentioned Bread manufacturing activities have been part of the past by various studies which have gone ahead and advanced through various technologies the new way of doing this is through machine learning. So, our study of literature is to understand what critical developments have happened in the recent past in this human activity recognition and how to we collaboratively use the approaches to build the solution for our business through the literature review. Below is the thought flow to continue the literature review from various scholars. Below is the objective and scope for the literature review for human activity recognition.

- What are the “*key developments*” which are “*readily available to use*” for the business?
- Literature review to scope in the “*intended topic*” to cover as per our project scope?
- Literature view for the “*additions*” which can be “*developed*” as per “*future scope*” which can be developed as part of our project?
- Is the future study feasible and flexible?
- What are the technological scopes which can help the business to move ahead?

- Literature review to understand what was out of scope and what can be scoped and reasons.

Human activity recognition: A review- 2014	
Authors	Ong Chin Ann; Lau Bee Theng
Study Take away	This paper was studied to understand the critical area of human activity recognition as part of the research study this is very critical study for various reasons however in this study the author has gone through 32 research sensing technologies like Cameras RGB, wearable devices, depth sensors. The author has discussed the merits and demerits of the sensing technologies. The observations out of the study were that RGB Cameras have less popularity when competed with the wearable devices and depth sensors in the HAR Research. (Ann & Theng, 2014)
Remarks	As part of our business model bread manufacturing factory, we have scope for both wearable devices and RGB Cameras as you see the cameras not so good compared to wearable devices and depth sensors, however the cost analysis for wearable devices or depth sensors will be one factor to assess the need of the hour

Semantic human activity recognition: A literature review – 2015	
Authors	Maryam Ziaeeffard ;Robert Bergevin
Study Take away	The study comprehends the method of activity recognition through semantic characteristics. This method provides the inherent features of the activities. Semantic makes recognition task more consistent when some actions look different because of the action executions. These actions in semantic features are recognised through the human body poses and pose let which are like attributes, scene content, related objects. The study presents methods to misusing of semantic features i.e., still images, videos data which even includes the other group of activities like people and human object interactions, atomic

	actions and group activities. This study even provides future potential applications approaches through semantic approach besides the future directions as well. (MaryamZiaeeefard & RobertBergevin, 2015)
Remarks	The semantic study for the bread manufacturing business seems to be very apt however as we see the analysis may really need lot of data sets, example , collecting and mixing raw material for bread, slicing, baking all the actions can be part of semantic human activities ;to study through machine learning predictable algorithms, however the scope of future industry to arrive the recognition will be sematic approach with machine learning algorithms

A review on applications of activity recognition systems with regard to performance and evaluation -2016	
Authors	Suneth Ranasinghe; Fadi Al Machot; Heinrich C Mayr
Study Take away	This research is conducted over <i>activity recognition system</i> which is scope for huge development in the activity recognition and scope of focus is all over innovations in the hardware filed and machine <i>learning algorithms. This research helps to decreasing the cost of monitoring and increase in the safety and efficiency.</i> This study majorly focuses over the applications which can recognize the systems and assist in live system like smart homes monitoring of healthcare applications, surveillance of both in and out activities and major tele immersion applications. In this study based on the behaviour the methods have been classified into multimodal sensor

	technology, visual and non-visual. The research provides the advantages and limitations of this methods and approaches. The representation is also about the datasets which can be used for such analysis and recognition. <i>The study summarizes with the comparison of the available and existing methods and methodologies</i> which when used in the real time environment allows to assist and guide to future approaches. (Ranasinghe, et al., 2016)
Remarks	This study seems good for the bread manufacturing industry however the approach itself says it needs lots of data sets to be examine in the real time world for the execution purpose, this approach of visual, non-visual and multimodal can be a future scope for the bread manufacturing factory

A Comparative Analysis of the Impact of Features on Human Activity Recognition with Smartphone Sensors -2017

Authors	Wesllen Sousa; Eduardo Souto; Jonatas Rodrigues; Pedro Sadarc; Roozbeh jalali; Khalil El Khatib
Study Take away	The study in this literature review gives understanding over the user's physical activities through smartphones and inertial sensors. With this there has been huge development and usage of this technology in the healthcare and transportation domains. Majority of these solutions have been aided and supported in the cloud technologies and communication as there is high risk related in both healthcare and transportation domains could computing provides the most accurate classification models. In an effort to solve problems which are related to the smartphone positioning (e.g. landscape) in the Experimental (user's) body, new types of classified features as part of orientation autonomous have ascended in the last years. With respect this topic or context, the paper study's and presents a

	<p>comparative study of all the features mapped in literature derived and inertial sensors.</p> <p>Experiments were conducted and using the two data bases which from 30 users the output of the orientation which is new couldn't be differentiate the user activities and through usage of the inertial sensors. The study even call outs that the features of the human activity recognition through various methodological based on smart phones (Sousa, et al., 2017)</p>
Remarks	<p>This study is an awesome approach for a healthcare or transportation industry however this may not be suitable for the manufacturing unit like bread were physical activities are through to be performed were the mobile usage or tracking can be source of distraction rather how we can use only assistance for login or admin privileges</p>

Comparison of offline and real time human activity results using machine learning techniques – 2018

Authors	Jozsef Suto, Stefan Oniga, Claudiu Lung & Ioan Orha
Study Take away	<p>Healthcare has been one of the biggest industries which has been very popularly known to use the human activity recognition in their business for example the <i>usage of the accelerometer and gyroscope sensor gives the raw data for the physical and functional activities of the experimental person.</i></p> <p>The prior studies have shown that the activities can be recognised through <i>machine learning with the help of pre-processing approach.</i> The studies done by the scientist in the recent past have been performed in labs and databases however real time experiments in the real time environment not performed, the objective of this study is to inspect the efficiency of the <i>previous machine learning approaches and methods in the real time environment with the help</i></p>

	<p><i>of self-learning, android based activity recognition applications.</i></p> <p>These applications have been precisely designed to study the theory's explained with most apt feature extraction and machine learning algorithms. Before inspecting and executing in the real time environment they have inspected the designs and applications. Different opportunities of deep and shallow methods. Output differences over the online and offline cases and reason for the differences in the output and possible solution to eliminate the differences. (Jozsef Suto, 2018)</p>
Remarks	<p><i>This study is very close to our business proposal infact this can be one of the approaches suggested for our business solutions in the Bread manufacturing industry as the usage of accelerometer ,gyroscope sensor can be used to collect the data monitor the employees in this manufacturing unit</i></p>

Human activity recognition through Sensor based Datasets :2018	
Authors	Emiro De-La-Hoz-Franco; Paola Ariza-Colpas; Javier Medina Quero; Macarena Espinilla
Study Take away	<p>This paper studies about the analysis of sensor-based data sets that can be used in HAR and identifying the appropriate data sets to analyse ARS and classify the techniques which can help create better results. The study was to help developing the quality of life in healthcare through HAR & ARS. To do this activity it is very much important to conduct the evaluation of the performances in recognizing the activities of day to day life and collect data to set the benchmarks though experimental scenarios. Techniques to predict HAR in specific context are the keys to progress in this research review. This study presents the sensor-based data to analyse ARS. These are done through various variables which are characterize publications through database, type quartile, origin of the country,</p>

	destination and year. Scient metrics is one of the key tool for identification of data sets among researchers. (De-La-Hoz-Franco, et al., 2018)
Remarks	This study makes more exiting to be followed however I see the sensor-based variables to define and collection of data sets and usage of latest technologies is definitely the scope for manufacturing units. For this lots and lots of data collection and accuracy test may need to be performed

A review of state-of-the-art techniques for abnormal human activity recognition - 2019

Authors	ChhaviDhimana; Dinesh Kumar; Vishwakarmab
Study Take away	The literature review of this authors provides human activity recognition for the abnormal activities, these activities are highly diverse with in itself due to anomaly, feature representation of anomaly, applications, and the data sets. Hence the feature design technique to tabulate are 2D and 3D AbHAR. The abnormal human activity recognition through a video with reference to the content or application like homeland security, Ambient Assistive Living, fall detection, crowd analysis using RGB, surveillance depth and skeletal evidence. The paper explains that the researchers have newly added datasets for AbHAR to analyse the complexities and for technique validations. (ChhaviDhimana, et al., 2019)
Remarks	This is the advance study over the abnormal behaviour and recognition through RGB camera, surveillance which is true development in the human activity recognition through machine learning we can only relay on the output once we accurate output, still this is in scope for the study for future manufacturing industry

Literature Review on Transfer Learning for Human Activity Recognition Using Mobile and Wearable Devices with Environmental Technology 2020

Authors	Netzahualcoyotl Hernandez, Jens Lundström, Jesus Favela, Ian McChesney & Bert Arnrich
Study Take away	<p>This research study is about the ARS -activity recognition system Using of the data from the mobile sensors and wearable devices which is available globally. This area of research has been very evolving and growing as author says that the model of identifying the human cognitive, social and physical actions, skills and patterns. this are especially reliant on the supervised machine learning techniques. This approach is little expensive due to the data collection is very diverse activity because of dynamic and interleaved of human behaviour. <i>The transfer learning approach is utilised to build a new model to with the previous data setting and patterns i.e. it can reuse the knowledge to recognise the activities performed by the users with the help of sensor technology and in varied environmental condition. As we adapt more internet of thing devices we may not escape from the mobile and wearable devices which intern are capable of handling more challenging human activity behaviours. The better data gets captured with the accurate labelling will moved into more better models to handle. Which is the advantage od transfer learning of activity recognition. The literature reviews and summarises the benefit of data set combining of both wearable and mobile devices through sensors with the help of transfer learning with the validation methods as per the reviews through the papers.</i></p> <p><i>Okali and Schabram methodology was most reviewed articles in which Inertial sensors such as gyroscopes and accelerometers are the most regularly used. For knowledge transfer Feature and instance illustration are established techniques. Typical application for the users is the Unsupervised learning and shallow methods.</i></p>

	Active learning is the scope of improvement in transfer learning techniques. (Netzahualcoyotl Hernandez, 2020)
Remarks	This study makes predictive learning through data sets the most advance for which we really have limitation to collect the data sets like in bread manufacturing unit- This is best study to go through machine learning however all the data sets need to be feed to the algorithms for the appropriate output and analysis

The above literature review was done from 2014 until 2020 however each case authors review were deeply studied to understand what approach can be considered for the Alpha bread manufacturing and what can be used or considered as future scope and study.

2.10 Conclusion:

Human activity recognition through machine learning is very critical and important study to understand or to analyse in the current evolving technologies era. As we see the easy of trends and technologies at the same time, we have even observed the complexity and complications of the human activity recognition as whole. the activity of recognition over the computer vision is an important problem and critical activity however there have been advances from the past to till date in business domains (healthcare, transportation, video surveillances and human vs computer interaction) and simultaneously in the methods and methodologies as well; however the challenges and problems in human activity recognition still exist in facing the real time scenarios besides there problem inherent interclass variation as well.

As we have seen the various studies and research of the various authors in the literature review to understand the criteria, we have gone through the business proposal of our study based on which we may classify the human activity recognition and conclude the literature study. In the preview the business in which are trying to adopt the human activity recognition is “Alpha bread manufacturing industry” in this industry we are trying to provide the solution through our literature

review what would be best solution for the reliant attitude of employees to drive to perform their roles efficiently and simultaneously how can we improve the productivity of the business through human activity recognition.

Before we take the reference to conclude the approach and futuristic plan just to understand what activities will be in scope for the Alpha bread manufacturing industry to get the best possible approach

Based on employee actions want to determine for they are productive or lethargic at work

Based on the activities scope and completion need to determine if that was the optimum time required to complete the activity

All are activities delivered time bound or is there latency due to non-competency and efficiency

Determining the efficiency of employees through the productive actions

Below are the sequence or actions to determine the approach collection of activities such as Walking downstairs, standing, sitting, laying, walking upstairs and walking.

Above-described actions are the key scope for the human activity recognition at high level to monitor by the system visually and provide the feedback for effective production

Various literature reviews were reviewed starting from analysing the human activity recognition by Anna and theng, 2014 to Transfer Learning for Human Activity Recognition Using Mobile (Netzahualcoyotl Hernandez, 2020) each literature review has its purpose and goals achieved however it has flaws in terms of either due to high technology which can return back to the data collection or cost or expense for the organization to bear it . based on the business criteria's we have selected below is the approach or method or literature study accept as solution and improve the efficiency of employees and increase the productivity in the business.

Some solutions have been considered for future plan and some problem areas where we may not be able to drive due to various issues

2.11 Literature review Futuristic scope to suite the business

- Literature review -human activity recognition by Anna and theng, 2014 : In this approach or model the scope is for wearable devices, depth sensors and RGB cameras. RGB cameras are not popular however wearable devices and depth sensors will be cost effective products for the business to choose
- Literature review -Human activity recognition through Sensor based Datasets :2018 (De-La-Hoz-Franco, et al., 2018)- in this study the collection of sensor based data sets and testing will be key area of concern for the bread manufacturing units
- Literature review A review of state-of-the-art techniques for abnormal human activity recognition -2019- this is advance study of abnormal behaviour not suitable for the bread industry as the solution will be pretty expensive and lot of data collection for the machine learning algorithms, this data set is collected for every now and then changes and abnormal activities
- Literature review -Literature Review on Transfer Learning for Human Activity Recognition Using Mobile and Wearable Devices with Environmental Technology 2020 (Netzahualcoyotl Hernandez, 2020) : this approach of data collection is devices dependent and all prediction base analysis through machine learning algorithm

Limitations

The human activities have been classified into three levels which includes the primitives, actions/activities, and the interactions. We further summarize them as classic and representative and classic approach for the activity classification and representation and benchmark datasets.

Classification methods play significant roles for advancing of HAR. The approaches are classified into template matching methods, generative models, discriminative models there 7 types of technique from the classic DTW to the latest deep learning which are summarized. For supporting human tracking techniques.

- Limitation 1- currently well-performed methods are hard to be execute in real time or applied to wearable devices, as they are subject to forced computing power.
- Limitation2 -Classification methods are anticipated to train recognition models really quick in even in real time.

- Limitation 3- Lower quality of input image and strike a steadiness among input data, algorithm effectiveness, and recognizing rate.
- Limitation4 -Recognition issues are resolved case by case for the benchmark datasets and the recognition methods. Uniting and splitting of datasets for algorithms (Shugang Zhang, 2017)

Best literature review as solution to the business:

Literature study - Comparison of offline and real time human activity results using machine learning techniques – 2018 (Jozsef Suto, 2018)

The above study as discussed in the literature gives the scope and can be used for monitoring the employees through accelerometer and gyroscope sensor which provides the raw data to analyse the action and understand the physical and functional activities of the user. Considering other approaches for the bread manufacturing based on cost and monitoring this approach can be the best solution however here there still a limitation and analysis of data collection which will be key area. Using the machine learning and prediction algorithms is the possible solution compare to other approaches. This theory has been very impressive and possible opportunity of executing in the real time environment with the helping of self-learning algorithms as discussed above the sequence of activities in the bread manufacturing business the solution looks good to implement. Below are some literature reviews which can be scoped in for future business research and development.

With the above review over the futuristic view, limitation and proposal I conclude my literature review for human recognition through machine learning.

Let us consider the implementation of this Human Activity Recognition as a project and drive this with the help of business modelling architecture, various UML diagrams, business strategies and project approaches.

To start with let us understand the business position of Alpha breads and relate our project and its benefits upon the implementation

Chapter 3

Business Modelling for Project

The business modelling for a project will have many steps which includes a Mission statement, Business Realization plan using a Balanced business scorecard, Swot analysis, Pestle analysis, Porters 5 forces, value chain analysis let us understand each of the aspects in briefly

3.1 Mission statement

“Monitor employee activities and improve their efficiency “

The mission statement highlights the mission of the organization which involves the major factors of the organization/project. in the context of the current project the mission of Alpha Bread is to monitor the activities and improve the efficiency of the employees, due to the careless activities of the production team, the effectiveness, and quality of the work process of Alpha Bread is going down. Improve the quality of the services and increase the productiveness of the product, are the other two major mission of the company. Making a human recognizing system using python, is also the mission of the company (Zhang *et al.* 2017, p. 1680). The company needs to stop careless activities such as sitting, standing, talking, and other activities during the production team, which affect the production rate of the company. It also affects the overall growth of the company.

3.2 Identification of IT project

The identification process of the IT project involves the development process of the project which involves various initiation of the product. The initiation of the product involves the prototype of the bread-making process which involves the basic functionality of the product. The prototype of the process highlights the overall working process of the company which involves all the basic equipment and the involvement of various methods and technologies. It project involves software related queries and language which helps to solve software issues and make an effective system using basic software (Cruz *et al.* 2020, p. 1). Here python and machine learning concept are used to make an effective system for the company which helps to reduce the error and risk part of the company. IT project helps to increase the quality of the system prototype which makes a strong

base for the company. It also involves the IT team who can make a software-based system to solve the issues of the company.

Scope Diagram

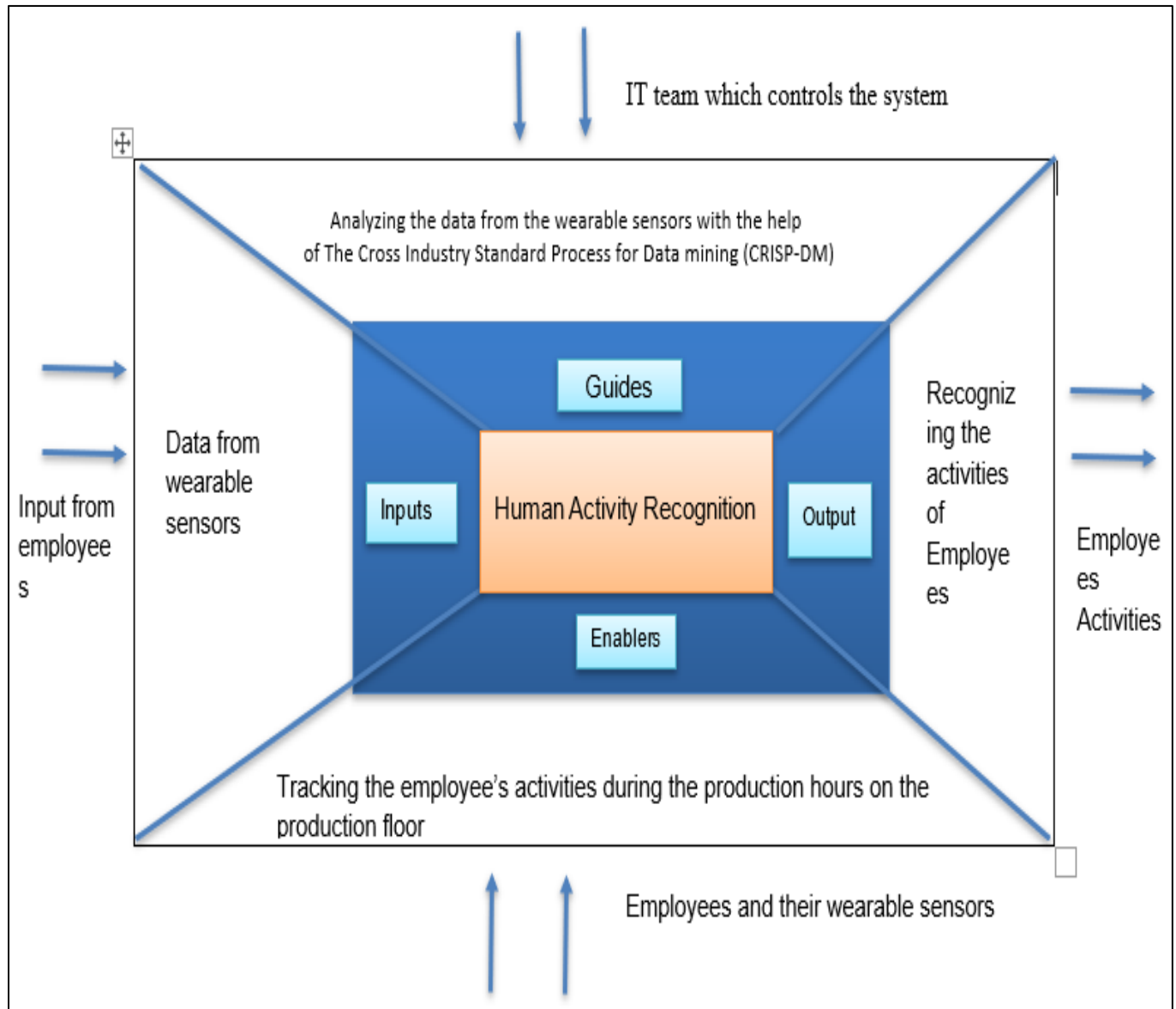


Figure 1 Scope diagram of HAR System

The above figure shows the scope diagram of the Alpha Bread company which involves various components and factors such as inputs, outputs, enablers, and guides. Those factors directly as well as indirectly affect the overall process of the project. Those factors have positive as well as

the negative impact on the overall growth of the company. According to the above figure, the input of the scope diagram of the project involves data from the wearable sensors of the employees. On the other hand, the output is the recognition of the activity. The enablers involve the employees and the wearable sensor devices. Guides involve CRISP-DM methodology which is utilized by the IT development team who are monitoring and analysing the data.

3.3 Producing requirements Catalogue

To start a project we would require catalogue which would have some functional and non-functional requirements

The requirement catalogue of the system involves all the basic components of the system such as non-functional and function components which helps directly and indirectly to increase the overall growth and productivity of Alpha Bread.

Functional requirement

The functional requirement of the system design of human activities involves basic components such as IT software involvement, the connection between various components of the system, and the involvement of various products. The functional requirement of the human activity detection system needs to involve innovation and technologies. To make the system more effective, a proper machine learning system and python code technology need to be applied here. Wearable sensors data analysis, activity analysis, time calculation are the functional requirement of the human activity detection process.

Non-functional requirement

The non-functional requirement involves employee's name, production floor, production team, date and time and other details of the system. Other details of the system involve the basic features of the human activity detection process and the effect of the system on the overall growth of Alpha Bread company. (Widodo *et al.* 2020, p. 012003). This involves sitting, standing, and the time gap between two activities which helps to analyse the fault of the system and careless activities of the employees.

Use case diagram

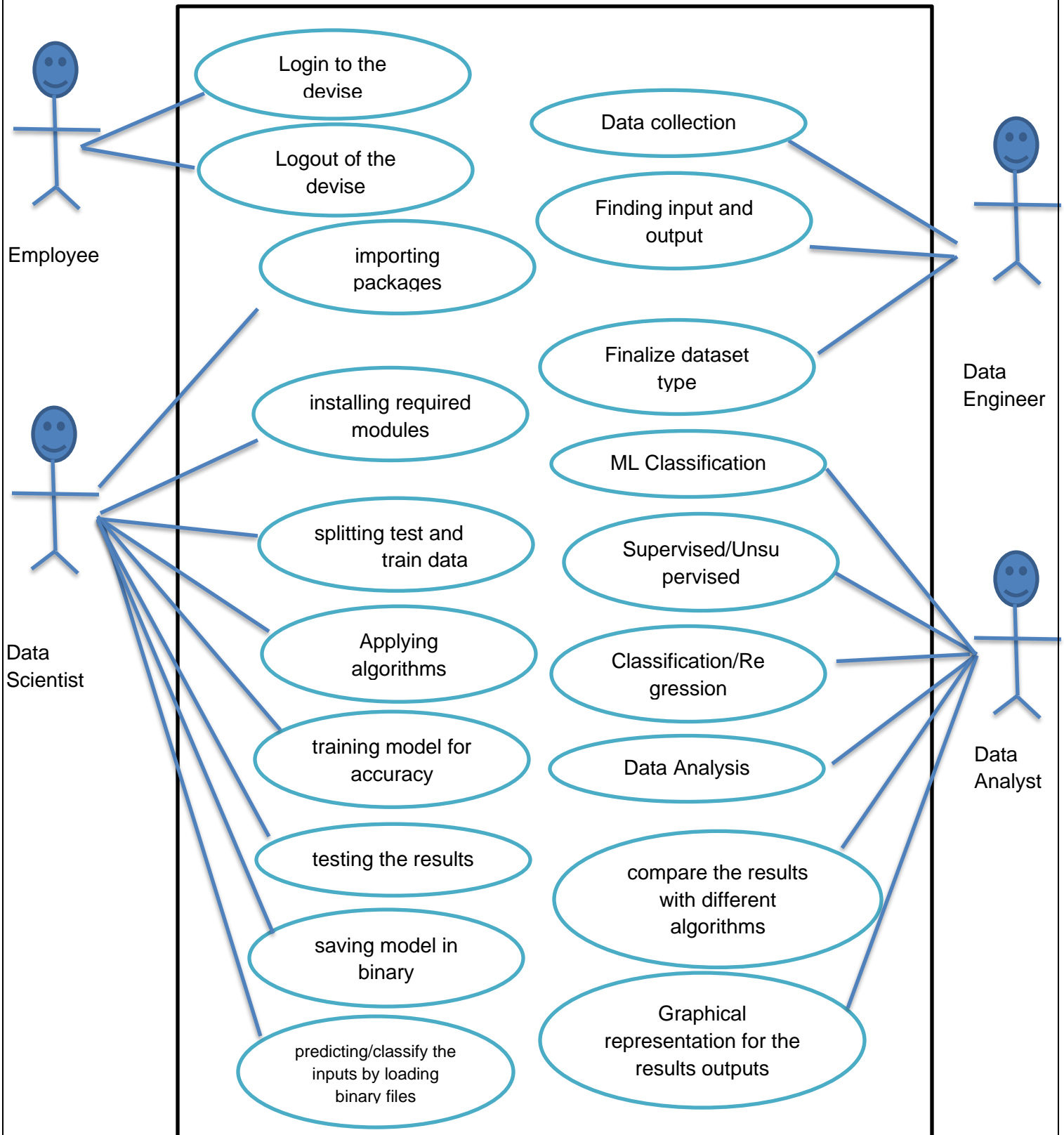


Figure 2 Use case diagram of HAR System

In the context of the project the use case diagram represents the actors and the activities performed by them. We have employees who has an activity of login and logout of the system and then we have Data engineer who performs the activity of data collection, who also identifies the input and output data and finalise the dataset type. Another actor who we have is Data analyst who does the Machine learning classification, differentiate the supervised and unsupervised data also performs the activity of classification and regression and does the data analysis. Data scientist is the another actor in this use case diagram who performs the maximum activities to start with he imports the packages, installs the required modules he adds the data if required and splits the train and test data post which he applies the algorithms and does the training for the model for accuracy then tests the results ones the results are our he will save the model in the binary format and predict the inputs by loading the binary files once all these activities are done either the Data scientist or the Data analyst will compare the different algorithms and represents the results in a graphical representation.

The whole process is done in real-time or this system is not useful. The system develops real-time data, and these are essential for stopping the employees to sit idle and announcements can be made for the employees to return to their stations (Chunget *al.* 2019, p. 1716). The system acquires the activities as developed in the diagram and the following the series of events an effective system can be developed.

These are the basic operations for functioning the data and these are essential for position monitoring of the system. The following aspect is developed for amplifying the signal and cancelling the noise of the signal for developing a high-definition image for analysis. The system then develops a human feature determination and velocity determination aspects (Zhenget *al.* 2018, p. 2146). The use case has been developed based on the functions generated in the motion determination and the employee monitoring systems. There is a login and logout method developed for the employees to assess their timing of arrival and departure on to the production floor. The system login and logout for managers and supervisors allow them to access the results developed to form the monitoring system and the log of a specific employee can also be detected using the system. This will help them take fair and necessary actions on the employee in positive or negative aspects. The different use cases are developed that lead to a specific activity to develop the images and develop a log based on the employee with the help of the system (Chenet *al.* 2017,

p. 3880). The system has also been effective in terms of speed measurement, human recognition, and action model development phenomena for developing motion sensing for the system of presence sensors.

Chapter 4

Project Initiation

Prince2 approach producing a PID document with a convincing project plan

To develop the project for supporting human recognition activity analysis concept for the employees of the Alpha Bread organization this section is going to focus on the project plan development factor. Prince 2 is a process-based approach and this approach helps to control the allover project work of an organization from the starting point to the ending point (Vaníčková, 2017, p.230). Application of the PRINCE 2 approach in the PID or the project initiation document helps to develop a stage-wise project development planning. The below figure helps to access the PID documentation for the staff idle time analysis IT project for the organization Alpha Bread:

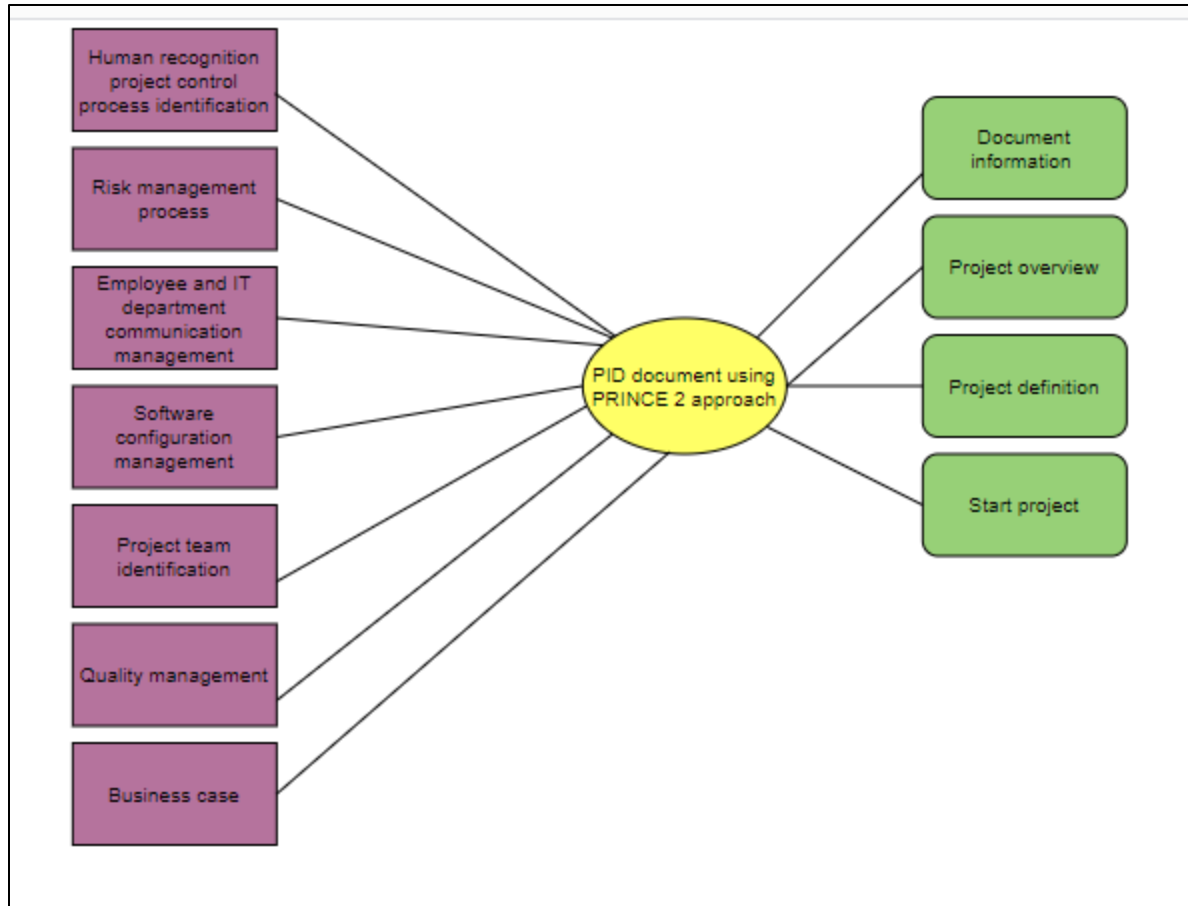


Figure 3 Project initiation plan document or PID using Prince2 approach

(Source: Self-created)

The project initiation document helps to cover several factors of the project work such as; project overview analysis, project definition development, project communication building, project initiation analysis and so on. Under the approach of PRINCE 2 the project initiation document helps to identify the project business case analysis to project control measure. Project development part of the Prince2 approach helps to obtain the human recognition IT project control measure for the organization by using the Prince2 tailoring process (Mousaei, and Javdani, 2018, p.9).

Project plan helps to integrate the PID document -based working factors by using strategic approach for project working structure development.

Table 1 Project plan

Project planning attribute	Description
Project requirement analysis	To find out the staff idle time under bread processing work of Alpha bread by using a human activity recognition system, a major requirement is to develop and utilize the employee working time dataset under the python language machine learning software.
Project team development	The IT team manager of the company has to be responsible to distribute the working role among the software solution team (Bijańska <i>et al.</i> 2016, p.128).
Quality management	Access and monitor the software working and configuration process to develop the results.
Risk management	Obtain the time, financial and IT knowledge related risk of the project.
Control process	Control the software team work and use the result for staff improvement work. Develop the project documentation.

(Source: Self-created)

4.1 Method description

This project development work uses the project activity-based method of Prince2 to develop the project initiation plan document or PID. The application of prince2 method considers for this project development by following the Price2 rule of Projects IN Control Environments. This method focuses on project initiation to project development work progression. This method helps to obtain the project management and control stage by emphasizing the project working structure. This method identifies communication and project risk management attributes and develops the project control measure steps facilities. Prince2 method follows project starting, project initiation, project management and project control steps under the PID plan (McGrath, and Whitty, 2020,

p.330). Project operational cost and risk analysis helps to select proper project management structure for the Alpha bread human activity recognition IT project work.

4.2 Requested total budget

At the time of project initiation the project development team has to develop the required project budget by accessing different project activity-related costing factors (Kwon, and Kang, 2019, p.90). After identification of the costing factors and required financial resources for each factor the requested total budget has to be sent to the financial manager of the bread making company Alpha Bread.

Table 2 Requested project budget

Project activity requirement	Cost
Project software and hardware, electricity requirement	£ 400
Remuneration of the software team	£ 2000
Quality analysis cost	£ 1000
Project documentation cost	£ 500
Project communication cost	£ 300
Total cost	£ 4200

(Source: Self-created)

To develop the human activity recognition project of the staff idle time analysis for the Alpha bread organization different costing factors has to be considered by the project development team such as; project software, hardware, electricity cost, quality management, communication management cost, documentation cost and so on. The above requested project budget table focuses on all those factors to propose the project cost budget for Alpha Bread.

4.3 Cost benefit analysis

The project cost benefit analysis focuses on the financial benefits analysis factor of a specific project work before starting a new project in the company. Application of the cost benefit analysis or the CBA technique helps to obtain the net benefit of the project work by accessing the total cost of the project (Annemaet *al.* 2017, p.109). In the case of Alpha bread this organization faces decrement in the bread making operation day by day due to lack of employee working involvement and efficiency. On that approach, the IT project required £ 4200 financial resources to develop the project activities. The Alpha bread organization increases their raw materials wasting amount day by day. This factor triggers lack of efficiency to manage financial resources of this organization. Assuming the raw materials wasting amount of this organization is £ 5000. Application of the staff idle time analysis project required £ 4200 cost. Application of this IT project helps to increase employee working efficiency by accessing and eliminating poor performing staff from the company. On that practice raw material waste cost became £ 0, for this organization after implementation of this project. The project benefit amount is $(£ 5000 - £ 4200) = £ 800$. This project helps to provide cost benefit for the bread making operation of the company Alpha bread by incorporating opportunity cost in this project.

Extra funding analysis

To support the e-mail communication and video conferencing-based communication the internet cost can be increased in future. On that purpose extra £ 200 can be required for each month. Also, software, hardware maintenance and electricity costs can be increased up to £ 500 per month. The value becomes £ 900. Project documentation cost can also increase with the developing code of practice of the organization. This estimated extra cost is £ 200. On that purpose the total extra funding requirement for this IT project is $(£ 200 + £ 500 + £ 200) = £ 900$ per month. To develop the project progress and maintain the project continuity assessment and incorporation of the forecasting or extra funding amount in the project budget is an essential approach of the project management team.

4.4 Project breakdown structure

Project breakdown structure helps to identify different working activities of the project and helps to categorize those activities under the four major sections of project development work. Major aim of developing the project breakdown structure is to develop the main tasks and subtasks of the

project work (Sutrisnaet *al.* 2018, p.393). The below figure is helps to access the project breakdown structure for the IT project work of the Alpha Bread.

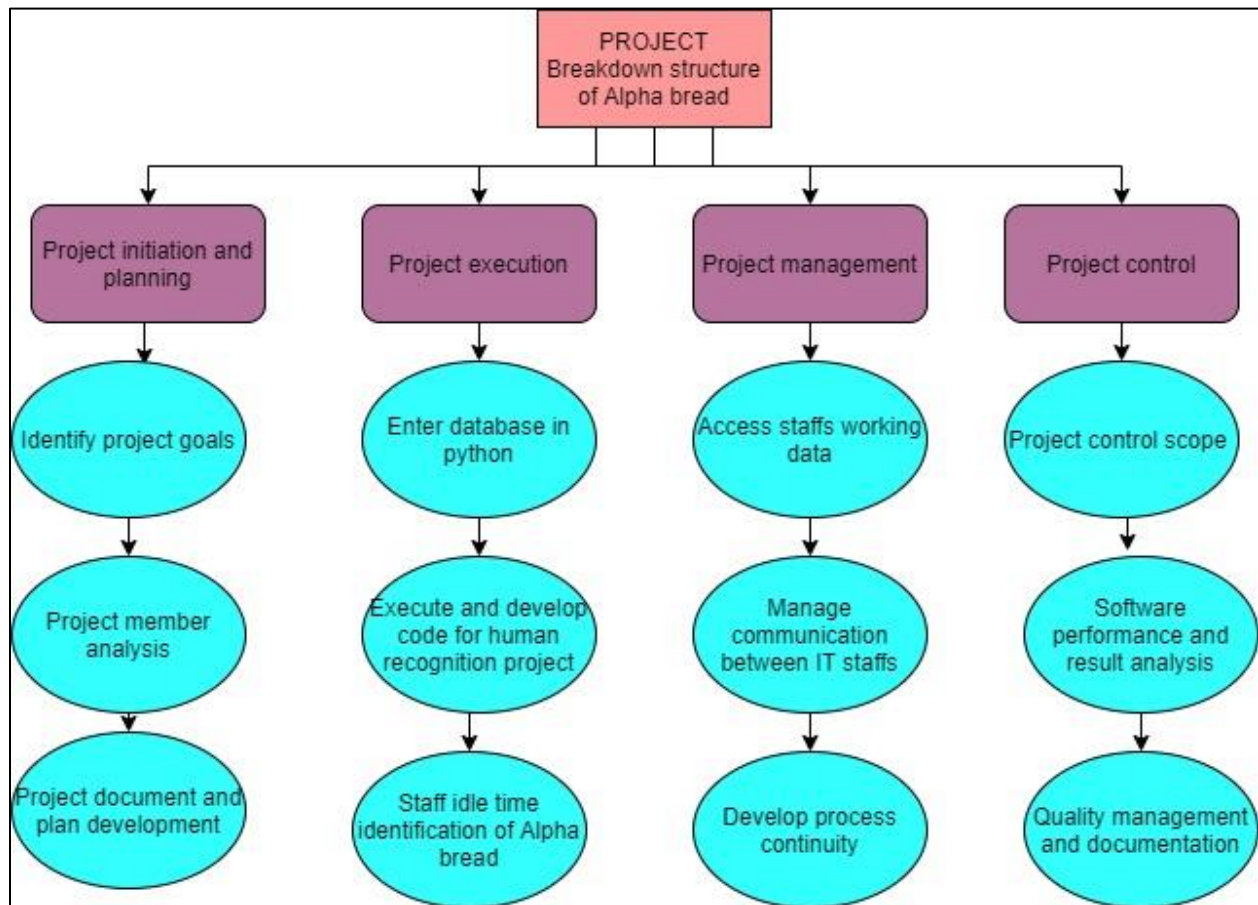


Figure 4 project breakdown structure

(Source: Self-created)

The project breakdown structure has three major parts when the first part of the breakdown structure contains the project goal and 100% work structure of the project. The second stage of the project breakdown structure follows project developing objectives and the third stage of the project develops the outputs or sub-tasks for the major four objectives of the project (Supriadiet *al.* 2017, p.823). The project management stage or task of the IT project follows staff working data analysis tasks, communication management and process continuity management tasks. The project control task has three subtasks such as; control scope identification, software work control and quality control work of the entire human activity recognition project for the Alpha bread organization.

Project activities and project flow analysis

Project activities and project flow analysis helps to develop a clear project plan to increase the quality and speed of the project work. Analysis of the project activities helps to obtain the project tasks and subtasks parts from the project initiation to project completion. Project activity is a milestone that helps to accumulate different work and tasks under the project planning strategy. The project flow analysis helps to access the task and subtasks working efficiency and its output quality as per the project planning design (Sutrisna *et al.* 2018, p.390).

The Application of project control activity helps to access the project flow, software data performance and overall quality analysis for this project work (Supriadi *et al.* 2017, p.830). The project flow analysis helps to assume the project completion and success details. Analysis of the project control activity helps to accumulate the performance details and quality measure of the python data and software-based staff idle time data output of the project. The control project captivity section is helps to build-up the project flow analysis of the IT project work and obtains the speed and quality of the project activity progression under this section.

4.4 Brief description of the entire process

The research methodology has demonstrated the execution process modeling and designing of the data driven human activity recognition system facilities within the organizational context. The evolution of suitable strategies to conduct the study has been considered in this regard. The implementation of projects based on the suitable project management facilities and critical pathways have been ensured in this regard. The development of suitable timeline based structure to meet the requirements of the infrastructure has been evaluated in this regard. The considerations of proper budget table and agile project development followed by risks analysis using the implemented coding using python language have been done. The engagements of data management and comparative analysis have been ensured in this regard to mitigate commenced risks using the available knowledge and analytical thinking.

4.5 Gantt chart and critical path analysis

Application of Gantt chart helps to identify the time requirement for the project execution activity completion purpose. The below bar structure format provides the project execution time requirement for the human activity recognition IT project work for the bread making organization Alpha Bread.

Table 3 Gantt Chart

SL no.	Weeks Works	1	2	3	4	5	6
1	Planning For human recognition activity project						
2	Data development and machine learning step						
3	Python program development						
4	Staff idle time analysis						
5	Consideration of project communication and quality						
6	Risk Analysis in the project						
7	Proper Mitigation of Risks						
8	Final launch of Project						

(Source: Self-created)

The critical path analysis of the project focuses on every activity of the project work and schedules the project work based on the project completion time. Application of the project activity with their respected time requirement helps to access the critical path analysis of the project

development work (Soe, and Htike, 2018, p.01027). The below figure helps to access the critical path analysis for the IT project work for the Alpha Bread:

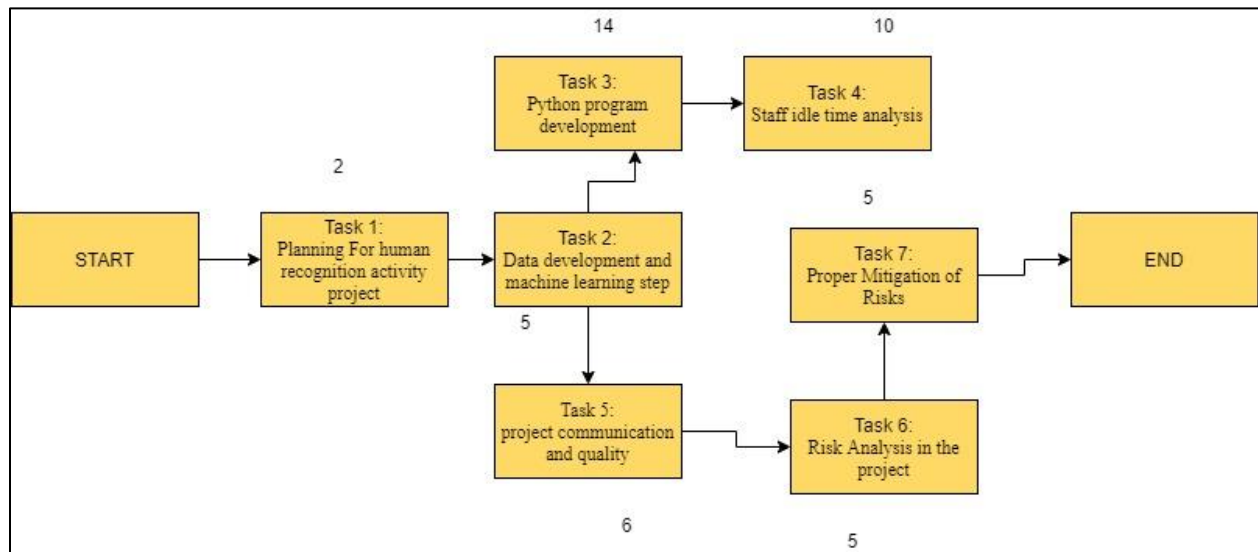


Figure 5 Critical path analysis of the project

(Source: Self-created)

The engagement of effective modeling and designing requirements has been initiated in this regard to optimize the performance of human activity monitoring and recognition systems. The involvement of effective business strategies and intervention of internet enabled infrastructure have been considered in this regard as a part of method in this study. The customer centric development of the process of facilities using underpinning knowledge of machine learning tools and technologies has been initiated in this regard. The agile project development process has been considered in this regard to guide the research. The relevance of the PID document, contingency plan and effective breakdown structure of the recognition system has been evaluated in this regard.

Chapter 5

Methodology and Design Human Activity Recognition System

In this chapter we will be understanding the design methodology and plan. For this data mining process several well-structured and robust approaches are identified. CRISP-DM Cross Industry Standard process form Data Mining is been chosen as the best method to conduct the current study as it is a well proven method historically (Piatetsky, 2014). There is some ideal sequence of events in this methodology and each phase of this will be dealt separately in this chapter these steps can be tracked back and forth and repeated to identify a better step suitable for the following stage. Understanding of the business is the done is the first stage, in this phase the perception of the project is perceived. The primary objectives of the study are the desired outputs of this phase. Once this stage is completed Data understanding phase will follow, in this stage the real data which would be used in the project is attained, all the elements of data will be inspected and described. Post this the data is still explored to generate any initial findings and their impact on the successive project stages. In this stage the quality of the data is accessed. The next phase is data preparation phase which is third phase here the repetitive data and the unnecessary data would be eliminated. Finally, the data elements that are manipulated are integrated creating an ultimate dataset which can be used in the next stages. Then comes the modelling stage. In the literature review section whatever modeling techniques are analyzed will be included here. In this stage different parameter settings are aligned with the estimates of the modeling methods. The later phase is very crucial stage which is the evaluation phase. With an evaluation criterion every individual model is evaluated with the prime focus. then models are assessed of their benefits and drawbacks (Wirth, R. and Hipp, J., 2000)

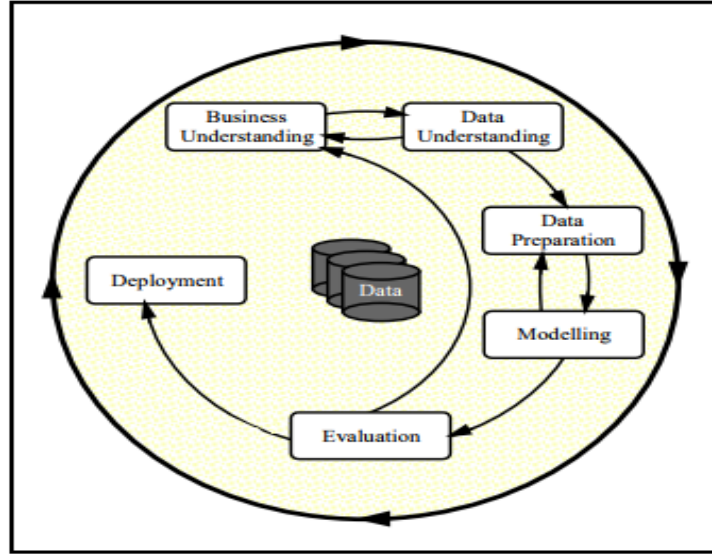


Figure 6 Different phases in CRISP-DM process Model

5.1 Hardware used

As this is report Is an academic submission, we have used a computer which is intel i3 7th generation 64 – bit model with 8 GB RAM, this may not have a high computing power hence the results may vary from the Realtime high performance computer used in the business

Each and every algorithm in the supervised machine learning developed on the HAR dataset will produce unique classification accuracy which would be greater than the benchmark SVM model Where the p value <0.05

The null hypothesis can be stated as

$$\left[\begin{aligned} & \left(\text{Accuracy (Decision Tree)} \right) \neq \left(\text{Accuracy(KNearestNeighbor)} \right) \neq \left(\text{Accuracy(NaiveBayes)} \right) \\ & \neq \left(\text{Accuracy(MultinomialLogisticRegression)} \right) \neq \left(\text{Accuracy(ArtificialNeuralNetwork)} \right) \end{aligned} \right] \\ > \left(\text{Accuracy (Base Support Vector Machine)} \right)$$

5.2 Understanding of DATA

The original data was taken from a collection six physical activities and the activities are walking, sitting, standing, climbing upstairs and downstairs, biking and jogging. These activities were performed by a group of 30 people within the age group of 19 to 48 under the supervision of the researchers

namely,

1. In the right pocket
2. In the left pocket of their trousers
3. one on the right leg using the belt clipper
4. One on the upper arm
5. One on the left wrist

Smart phones were used in the first three positions and smart watches were used for the fourth and fifth positions the phone used in this experiment was Samsung Galaxy SII (i9100) smartphone. The data was recorded for all the five positions at the same time and the data was collected at the rate of 50 samples per minute this sample size would be enough for the deduction of human activity and the data was collected from an app by adding linear acceleration sensor the data was collected from an accelerometer, a gyroscope magnetometer and a liner acceleration sensor. The team has captured three axial linear acceleration and 3 axial angular velocity at a constant rate of 50Hzs

The same dataset was uploaded to UCI repository which is a free source and we have downloaded the dataset from them

Accelerometer (Ax = x-axis, Ay = y-axis, Az= Z-~~ax~~~~is~~)

Linear Acceleration Sensor (Lx = x-axis, Ly = y axis, Lz= Z-~~ax~~~~is~~)

Gyroscope (Gx = x-axis, Gy = y-axis, Gz= Z-~~ax~~~~is~~)

Magnetometer (Mx = x-axis, My = y-axis, Mz= Z-~~ax~~~~is~~)

5.3 DATA Preparation

This is one of the most important stage and it consumes the maximum time of the of data mining process. The quality of the data is determined in this process and it extracts the insights that would in return influence the insights quality. In this process the data is manipulated and made appropriate for the subsequent phases of the process. The cleaned data from this process will eliminate any inconsistency, noisy and incomplete data

Below mentioned are the steps involved in the data preparation

5.4 Cleaning of the DATA

Evaluating the missing data from the complete dataset is the first and foremost task. N/A or 999 or not applicable is denoted or kept for the missing value in the dataset as an alternative for a blank field. Missing values is very regular incidence, and it will have a substantial result on the insights and features strained from it. Upon the identification of missing value is done then we need to know the mechanism behind it. The next step is to analyze the missing data types,

Missing data is of 3 types

1. Missing not at Random- MNAR
2. Missing at Random – MAR
3. Missing completely at Random

At the beginning of the experiment a suitable resolution must be developed and emphasized. Next comes the outlier detection and analysis. Any data point which is at an extensive detachment from the rest of the data points is called outliers (Grubbs, 2012). Error in data collecting method or in coding will give rise to an outlier, however there might be a possibility that the outlier would be an honest likelihood and it could be quite understanding, but if a data point is just an outlier and if it is suggestively far from the usual curve, here histograms and box plots could assist analyse in this situation. when spotted, it can be exposed to swamping, masking or any other formal outlier tests as Grubbs test. Before proceeding to next step with the experiment the complete outlier details, test outcomes must be noted (Croarkin, 2006). Handling noisy data is the next stage of the data cleaning process. (Zhu, 2004) according to him noise is an unavoidable problem in a data mining problem. Class noise and attribute noise are the two types of noises. When the target

variable is wrongly labelled with an even contradictory examples or with a different class a class noise will occur, a fault in the values of features is called Attribute noise. This can be an unknown symbol or an inaccurate value in place of an incomplete or actual value. Applying a noise filter is an appropriate option to avoid noisy data. It will also help to collect binning data or only eligible values into several bins and thereby averaging the values to common bin. An alternative for this solution could be utilizing a modelling approaches that could be resistant to noisy data as fuzzy decision trees (Michael D. Blechner, 2005). Another common problem with data is inconsistency. There might inconsistencies in data such as date formats. Naming conventions or data codes. Incorrect data representation could cause the data inconsistency. This inconsistency can be caused while creating the original data source and caused if inconsistent codes are used. Examining the data for consecutive rules, unique rules and null rules is the best solution. Eliminating data redundancy is the final stage of data cleaning process. If you have a huge amount of redundant data, it may divert the model from focusing the important observations. It may also sluggish down the whole process. The process of performing correlation between the features which would provide the relation among the features can detect redundancy. Correlation coefficient is the most popular correlation metric.

A strong relation is something where a correlation coefficient is closer to a negative or positive, where one of the features can be represented by another and feature and the other can be dropped. The process of combining data from multiple sources is called Data integration. This process can help in achieving an improved accuracy of the final prediction results as well as increase the speed of the experiment. Identifying the features that need to be combined and matched is a crucial process while performing integration. It is very important to understand the structure of such features very well along with their functional rules, referential constraints and dependencies which must match for both the integrated features and their datasets, if they fail to have similarities, the data and its conditions must be transformed to make required changes. (Vellampalli, 2017) During the data cleaning stage Redundancy tests need to be to evaluate and improve the performance for a better integration. The conflict between numerous values must be detected by the end of this stage that are caused due to the inequalities in encoding, scaling or representation and thus offer a suitable resolution.

5.5 Transformation of DATA

The available data is transformed from the present form to another form by applying a mathematical function, so the resulting value is more appropriate for the subsequent stages of the data mining process. There are multiple methods for transformation of data.

Aggregation

It is one of the techniques where summarization is applied to the data. A data cube is created by utilizing this technique of transformation. This is a 3-D range of values which are aggregated from distinct abstraction and aggregation patterns. Aggregation can aid when individual observation can be better characterized as a single aggregated value instead of multiple values partly explaining the same feature.

Normalization

It is another technique for ensuring the data falls in a static range of values. It is performed since the measurement of unit can have a substantial effect on the final result. Which would avoid dependency on the measuring unit and individual value in a feature is scaled using the standard deviation – mean or minimum – maximum of all the values in the feature. Normalization gives same amount of importance is given to each feature to avoid the algorithm getting influenced by higher values of feature.

Data reduction

It is a process of reducing the data and increasing the storage capacity, it will reduce the cost of data storage and increases the efficiency. There are multiple ways to reduce the data one of the ways is to identify the duplicate values and eliminate them, data duplication is also called as dedupe another way is to compress the data where the data is compressed in terms of bits and make some additional space in the storage capacity. The third way is single instance storage which identifies the files that are used multiple times and remove them and store only one copy of that.

Dimensionality

It is the number of input variables or features for a dataset. The techniques of reducing the number of input variables is Dimensionality reduction. The predictive modelling task will become

challenging if the number of input variables are high. Dimensionality reduction is used for the data visualization these techniques are also used for simplifying the regression or classification of the dataset. Feature selection methods, Autoencoder methods, manifold learning, Matrix Factorization are some of the Dimensionality reduction techniques

Principal component analysis (PCA)

It is a technique which is similar to Dimensionality reduction, this technique is helping to reduce large data set of variables in to smaller ones without losing its main content. Accuracy will be at risk if the number of variables of a data set is reduced. It will be easy to visualize and explore and analyse the data for machine learning algorithms using principal component analysis without extraneous variables to process. There are 3 steps involved in PCA.

1. Standardization
2. Covariance matrix computation
3. Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components

5.6 Confusion Matrix :-

This is the most persuasive tool used for predictive analysis in machine learning. Confusion matrix is used to check the performance of classification-based Machine Learning Model. Confusion matrix will show how a machine classifier is performed.

1. A confusion matrix presents the ways in which a classification model becomes confused while making predictions.”
2. A good matrix (model) will have large values across the diagonal and small values off the diagonal.
3. Measuring a confusion matrix provides better insight in particulars of is our classification model is getting correct and what types of errors it is creating.

True Positive, True Negative, False Positive and False Negative

Confusion matrix is a performance measuring method for all machine learning problems based on classification. Confusion matrix is a table which has 4 separate combinations of actual and predicted values. Positive and negative values are the predicted values, and the True and False values are the actual values, these four form building blocks of the Confusion matrix

n=165		Predicted: NO	Predicted: YES	
Actual: NO		TN = 50	FP = 10	60
Actual: YES		FN = 5	TP = 100	105
		55	110	

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 7 Confusion Matrix

Chapter 6

Implementation

6.1 Load DATASET

Below are the steps involved in loading the dataset

1. Download the dataset from sensor data from UCI Machine learning repository
2. Create new environment in the Anaconda Prompt using command
conda create -n HAR python=3.8, HAR is the name of the environment created and activate my environment by typing **conda activate HAR**
3. Go to path of dataset and set Path to the folder created for Documents:
cd C:\Users\Lenovo\Documents\MSC_Project\HAR
go to the working location or directory launch jupyter notebook
4. Create a new file name by clicking python3 then a. pynb will be created

6.2 Splitting of the data

Data splitting is a technique used for splitting of the data in a machine learning project there are generally 3 types of data splits and they are Training set, Testing set, and cross validation set among the three the training set is the largest set, each algorithm will be trained on the training data set and evaluated against the test dataset. The ratio of these two datasets would be 80 and 20 or 70 and 30 respectively for training and test data sets. (Liberty et al., 2016)

6.3 Build the model

This is the stage where the machine learning algorithm is used to include the predictive model. As mentioned, is our literature review and based on our reach there are multiple classification algorithms which can be utilized in machine learning and HAR domain. All these available algorithms are compared, and we might definitely find some gaps in them.

Below mentioned are some of the algorithms that are tested

6.4 Exploratory data Analysis (EDA)

This is a supplement product of inferential statistics, this analysis includes many formulas and rules, EDA looks the dataset from various angles and describes the outcomes and summarizes it.

Data Analysis is the probability and statistics to find the trends in the data set. Historical data can be seen and analysed with the help of analytical tools. Information can be drilled down and represented as facts, metrics and figures

There are many steps involved in the EDA namely

Missing value handling

Duplicates removing

Outlier detection

Scaling and normalizing

Bivariate Analysis

Encoding categorical variables

1. Reading the data and importing necessary libraries to load the data

2. describe the data from train and test DataFrame as

Table 4 Table 4 x_train dataframe

1	# Describing X_train dataframe											
2	X_train.describe()											
	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
count	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000	7352.000000
mean	0.274488	-0.017695	-0.109141	-0.605438	-0.510938	-0.604754	-0.630512	-0.526907	-0.606150	-0.468604	-0.306043	-0.306043
std	0.070261	0.040811	0.056635	0.448734	0.502645	0.418687	0.424073	0.485942	0.414122	0.544547	0.282243	0.282243
min	-1.000000	-1.000000	-1.000000	-1.000000	-0.999873	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000
25%	0.262975	-0.024863	-0.120993	-0.992754	-0.978129	-0.980233	-0.993591	-0.978162	-0.980251	-0.936219	-0.563561	-0.563561
50%	0.277193	-0.017219	-0.108676	-0.946196	-0.851897	-0.859365	-0.950709	-0.857328	-0.857143	-0.881637	-0.479677	-0.479677
75%	0.288461	-0.010783	-0.097794	-0.242813	-0.034231	-0.262415	-0.292680	-0.066701	-0.265671	-0.017129	-0.065364	-0.065364
max	1.000000	1.000000	1.000000	1.000000	0.916238	1.000000	1.000000	0.967664	1.000000	1.000000	1.000000	1.000000

Table 5: *y_train* Dataframe

1	# Describing y_train dataframe
2	y_train.describe()
0	
count	7352.000000
mean	3.643362
std	1.744802
min	1.000000
25%	2.000000
50%	4.000000
75%	5.000000
max	6.000000

Visualise the HeatMap

```
1 import matplotlib.pyplot as plt
2 import seaborn as sns
3 auto_data = X_train
4 corr = auto_data.corr()
5 sns.heatmap(corr,
6             xticklabels=corr.columns,
7             yticklabels=corr.columns)
8
9 plt.show()
```

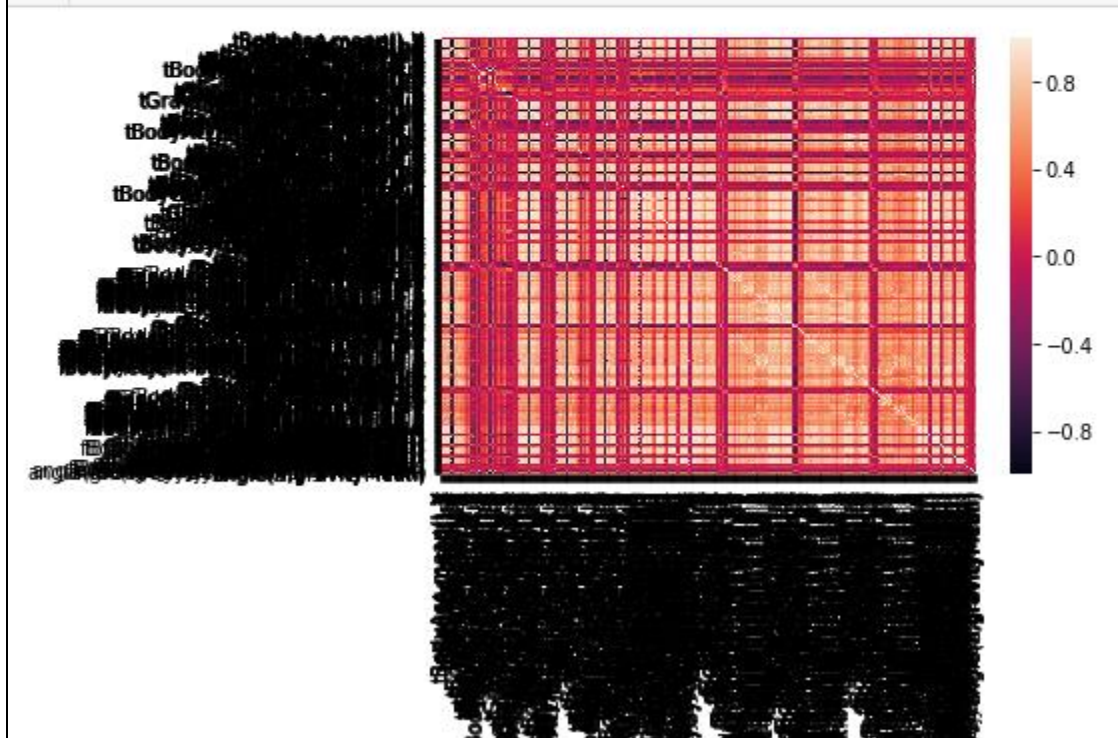


Figure 8 Heat Map

To know how data looks let's plot bar graph

```

13 fig.subtitle('The features are already normalized and range is [-1 to 1]. This shows normal distribution of data.', fontsize=12)
14
15 axs[1].hist(X_train.max(),bins=10,facecolor='g', alpha=0.75)
16 axs[1].set_xlabel('Maximum of each column')
17 #axs[1].set_title('')
18 axs[1].set_ylabel('Count')
19
20 plt.show()
21

```

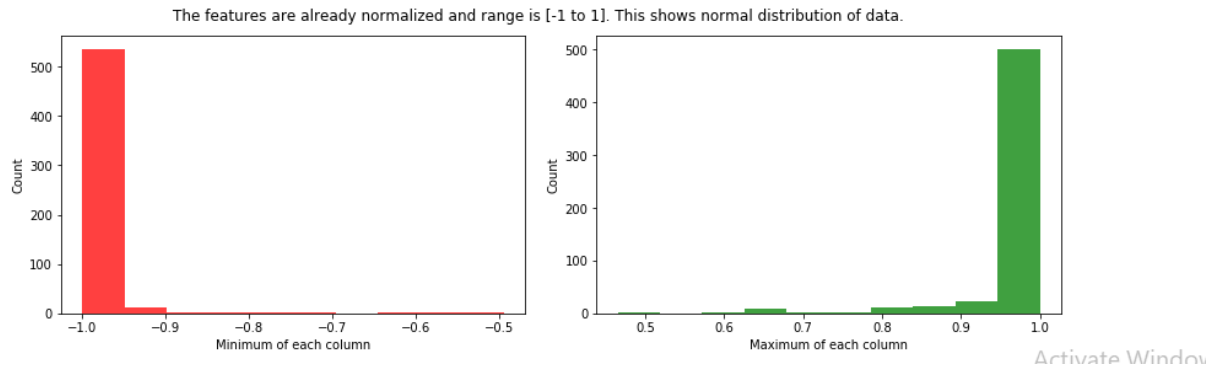


Figure 9 Bar Graph

Plotting the features

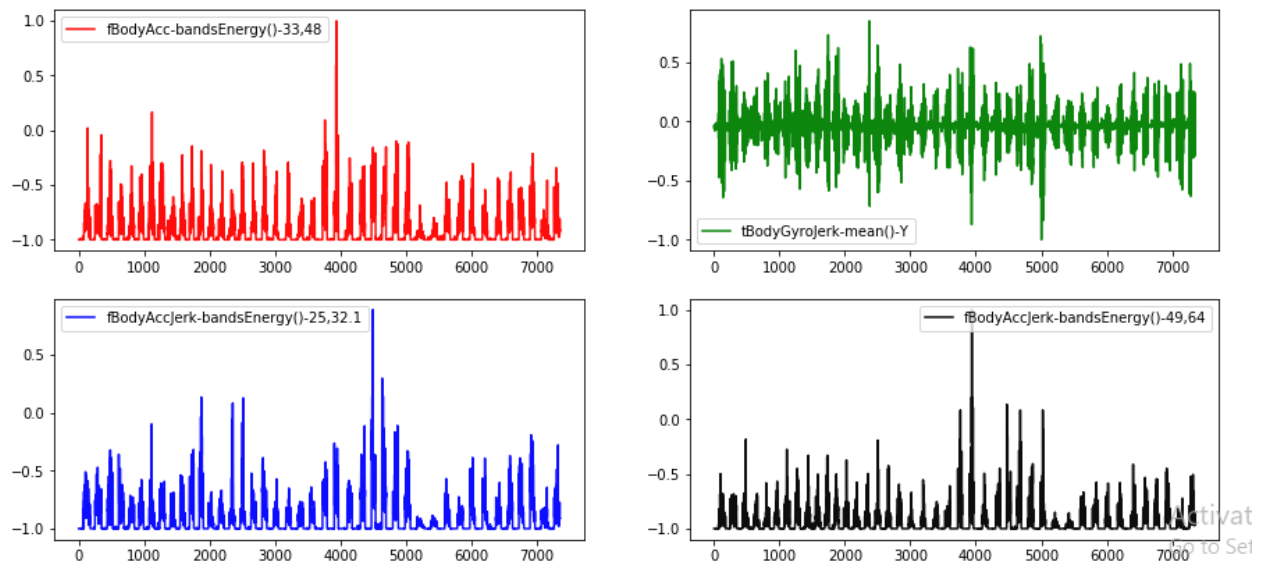


Figure 10 Line Graph

Visualising the data frame for both Activity and no Activity

Visualizing both the dataframes

1	activityDataframe.head(2)													
	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	0.282022	-0.037696	-0.134897	-0.328280	-0.137153	-0.189086	-0.404875	-0.176791	-0.199474	-0.154731	-0.203759	-0.130752	0.11207	0.11207
1	0.255841	-0.064550	-0.095186	-0.229207	0.016506	-0.260311	-0.276764	0.009066	-0.287367	-0.154731	-0.173983	-0.156627	0.11207	0.11207
1	noActivityDataframe.head(2)													
	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	0.852947
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	0.849308

Figure 11 DataFrames

Smoothing the data with Activity Dataframe and comparing with NoActivity Dataframe and creating Line graph for it

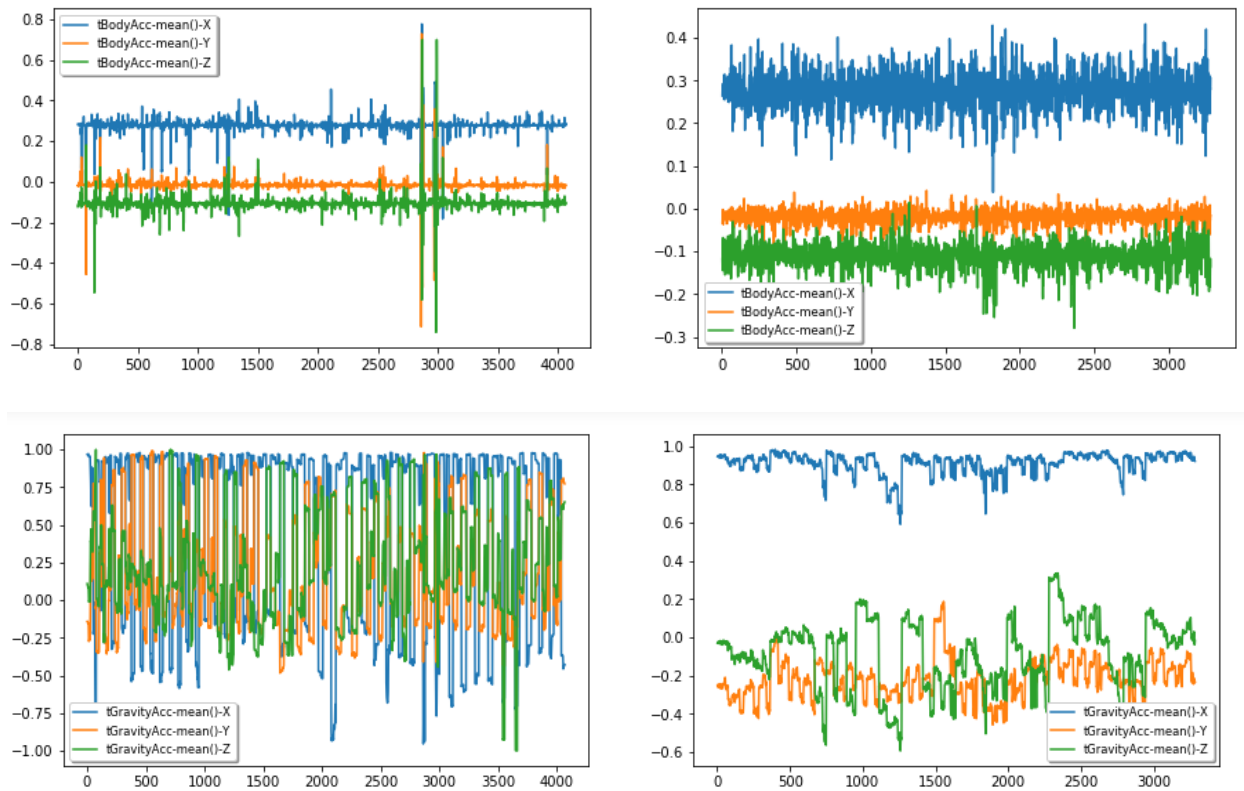
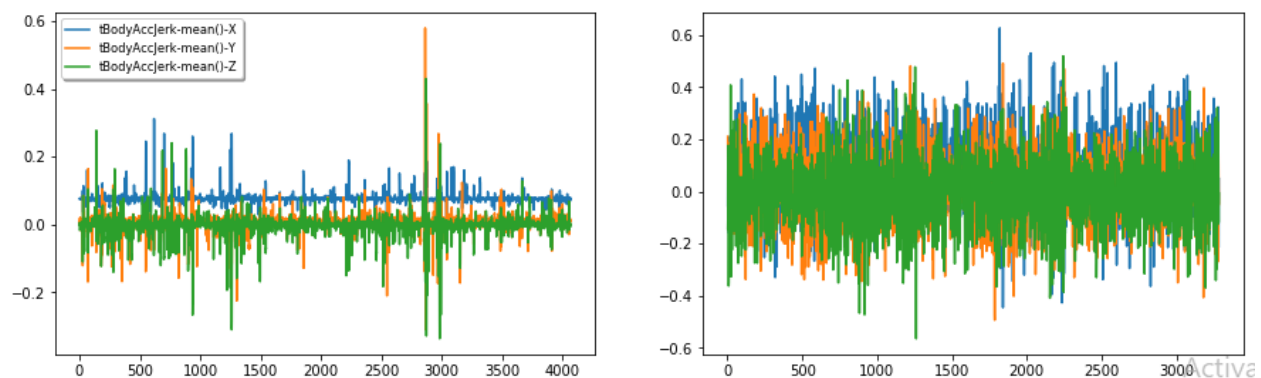


Figure 12 Line Graph



To check whether class in y feature are equally distributed



Figure 13 Histogram of features

Make a list of features to randomly sample of 561 features

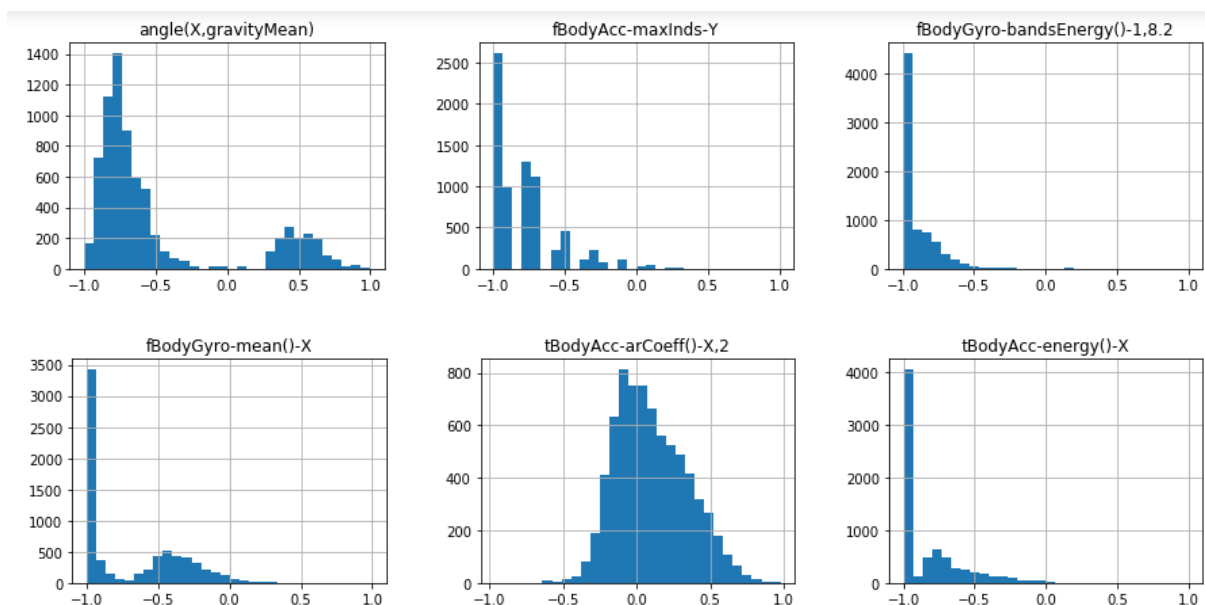
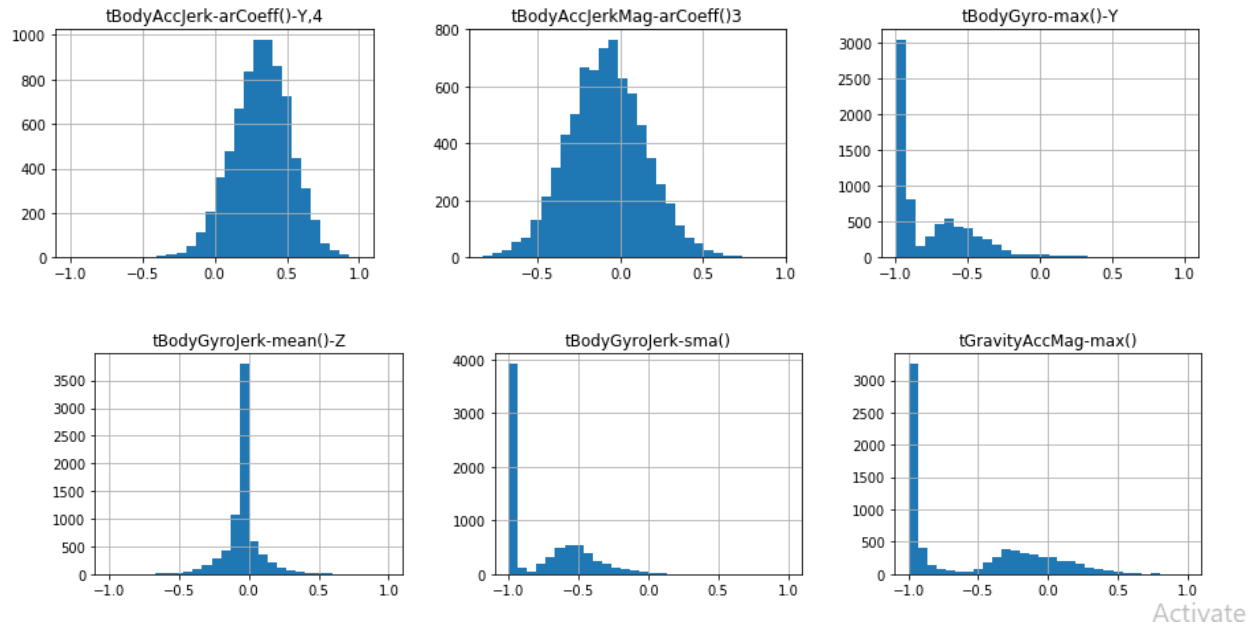


Figure 14 All list of Features



Model Building

1.logistic Regression Classifier

This is a classification algorithm; this is utilized when the value of the target variable is categorical in nature. This classifier is used when the output is in binary format and the outcome is either 0 or 1. Unlike other regressions tasks this logistic regression classifier have discrete categories. This regression classifier is used to solve the classification problem.

Below mention is the python code for the logistic regression classifier along with its results for HAR

Table 6 Logistic Regression Classifier

```
c:\python36\lib\site-packages\sklearn\linear_model\logistic.py:432: FutureWarning:
0.22. Specify a solver to silence this warning.
FutureWarning)
c:\python36\lib\site-packages\sklearn\utils\validation.py:724: DataConversionWarning:
array was expected. Please change the shape of y to (n_samples, ), for example:
y = column_or_1d(y, warn=True)
c:\python36\lib\site-packages\sklearn\linear_model\logistic.py:469: FutureWarning:
'0' in 0.22. Specify the multi_class option to silence this warning.
"this warning.", FutureWarning)
```

Accuracy of logistic regression classifier on test set: 0.96				
	precision	recall	f1-score	support
1	0.94	1.00	0.97	496
2	0.97	0.95	0.96	471
3	1.00	0.97	0.98	420
4	0.97	0.88	0.92	491
5	0.90	0.97	0.94	532
6	1.00	1.00	1.00	537
accuracy			0.96	2947
macro avg	0.96	0.96	0.96	2947
weighted avg	0.96	0.96	0.96	2947

k-Nearest Neighbour

KNN is a model which classifies data points based on the points that are most similar to it. This model makes an educated guess. It is very easy to use and quick in calculating time. It does not make any assumptions about the data; however the accuracy is depended on the quality of data, in this model we should find an optimal k value which is a nearest neighbours

[K-Nearest Neighbours \(KNN\) Algorithm for Machine Learning | by Madison Schott | Capital One Tech | Medium](#)

Below mentioned is the python code for KNN along with the accuracy for HAR

Table 7 Naive Bayes Classifier

```
1 from sklearn.neighbors import KNeighborsClassifier
2
3 xx=np.array(X_train)
4 yy=np.array(y_train)
5
6 neigh = KNeighborsClassifier(n_neighbors=3)
7 neigh.fit(xx, yy)
8 KNeighborsClassifier(...)
9
10
11 result=neigh.predict(np.array(X_test))
12
13 from sklearn.metrics import accuracy_score
14 z=accuracy_score(y_test,result)
15 z
```

c:\python36\lib\site-packages\ipykernel_launcher.py:7: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

import sys

0.8907363420427553

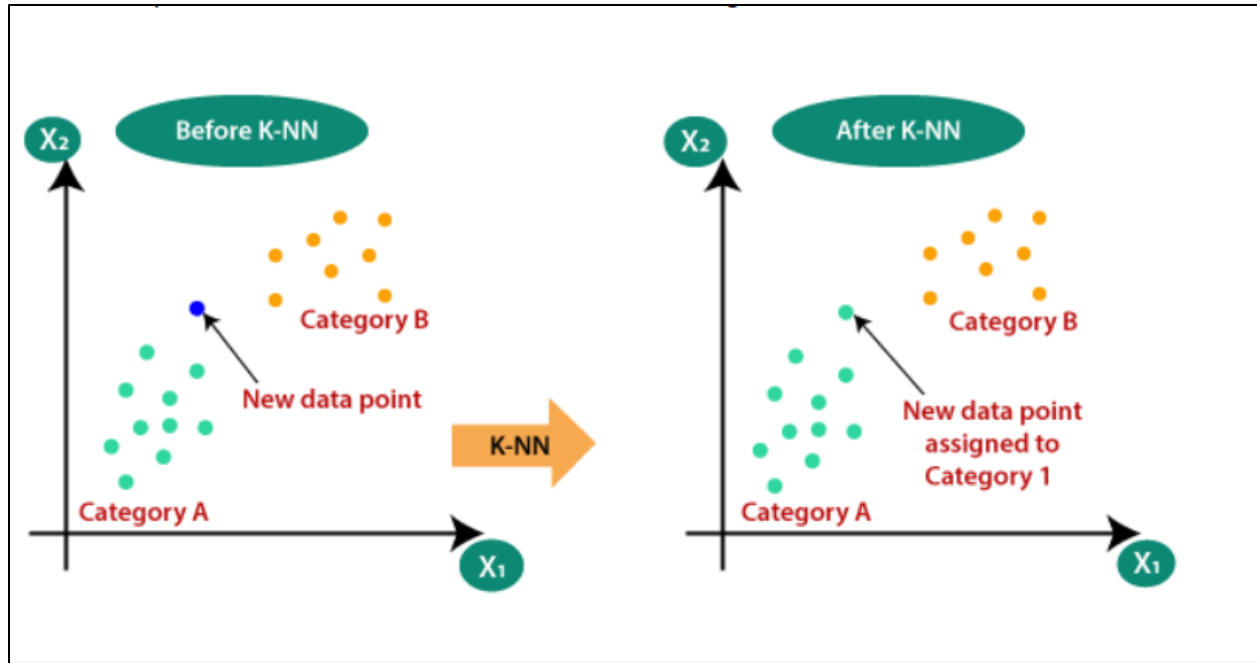


Figure 15 KNN Classifier

SVM Classifier

This classifier is a supervised machine learning model which can be used as classification algorithm for two group classification problems

Below mentioned is the python code using SVM algorithm along with the accuracy result

SVM

```
1 from sklearn import svm
2 classifier = svm.SVC()
3 classifier.fit(X_train, y_train)
4
5 results = classifier.predict(X_test)
6
7 from sklearn.metrics import accuracy_score
8 z=accuracy_score(y_test,results)
9 z
```

```
c:\python36\lib\site-packages\sklearn\utils\validation.py:724: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
c:\python36\lib\site-packages\sklearn\svm\base.py:193: FutureWarning: The default value of gamma will change from 'auto' to 'scale' in version 0.22 to account better for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.
"avoid this warning.", FutureWarning)
```

```
0.9402782490668476
```

Figure 16 SVM Classifier

Decision Tree

This is a modelling tool which has an application spanning a number different areas. In decision tree always there will be two different conditions and this algorithm approach identifies the split data. This is a form of supervised learning which is most commonly used. Decision tree is used for both classification and regression tasks.

Below mentioned is the python code based on Decision tree

Random Forest

Table 8 Random Forest Classifier

<pre>2 from sklearn.metrics import classification_report 3 import time 4 5 clf = RandomForestClassifier(n_estimators=20) 6 start = time.time() 7 clf = clf.fit(X_train, y_train) 8 print('Random forest took %.2f seconds' % ((time.time() - start))) 9 y_predict = clf.predict(X_test) 10 11 print(classification_report(y_test, y_predict)) 12 13</pre>					
c:\python36\lib\site-packages\ipykernel_launcher.py:7: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel(). import sys					
Random forest took 2.34 seconds					
	precision	recall	f1-score	support	
1	0.85	0.96	0.90	496	
2	0.87	0.85	0.86	471	
3	0.95	0.84	0.89	420	
4	0.90	0.88	0.89	491	
5	0.89	0.91	0.90	532	
6	1.00	1.00	1.00	537	
accuracy			0.91	2947	
macro avg	0.91	0.90	0.91	2947	
weighted avg	0.91	0.91	0.91	2947	

Activate Windows
Go to Settings to activate Windows.

K-Fold Cross Validation

K-Fold CV is where a given data set is split into a **K** number of sections/folds where each fold is used as a testing set at some point. Let's take the scenario of 5-Fold cross validation(K=5). Here, the data set is split into 5 folds. In the first iteration, the first fold is used to test the model and the rest are used to train the model. In the second iteration, 2nd fold is used as the testing set while the rest serve as the training set. This process is repeated until each fold of the 5 folds have been used as the testing set.

K-Fold Cross Validation. Evaluating a Machine Learning model can... | by Krishni | Data Driven Investor | Medium

Table 9 K-Fold Cross Validation Classifier

<pre>----- Cross Validation ----- (7352, 561) (7352,) (7352, 1) c:\python36\lib\site-packages\ipykernel_launcher.py:6: FutureWarning: Method .as_matrix will be removed in a future version. Use .values instead. c:\python36\lib\site-packages\ipykernel_launcher.py:7: FutureWarning: Method .as_matrix will be removed in a future version. Use .values instead. import sys c:\python36\lib\site-packages\ipykernel_launcher.py:20: FutureWarning: Method .as_matrix will be removed in a future version. Use .values instead. Accuracy: 0.93 (+/- 0.08)</pre>
<pre>1 import sklearn 2 scores = sklearn.model_selection.cross_val_score(3 clf, X_train_np, y_train_np, cv=10) 4 5 print('Accuracy: %.2f (+/- %.2f)' % (scores.mean(), scores.std() * 2))</pre>
<pre>Accuracy: 0.92 (+/- 0.08)</pre>

Expected Verses Predicted curve

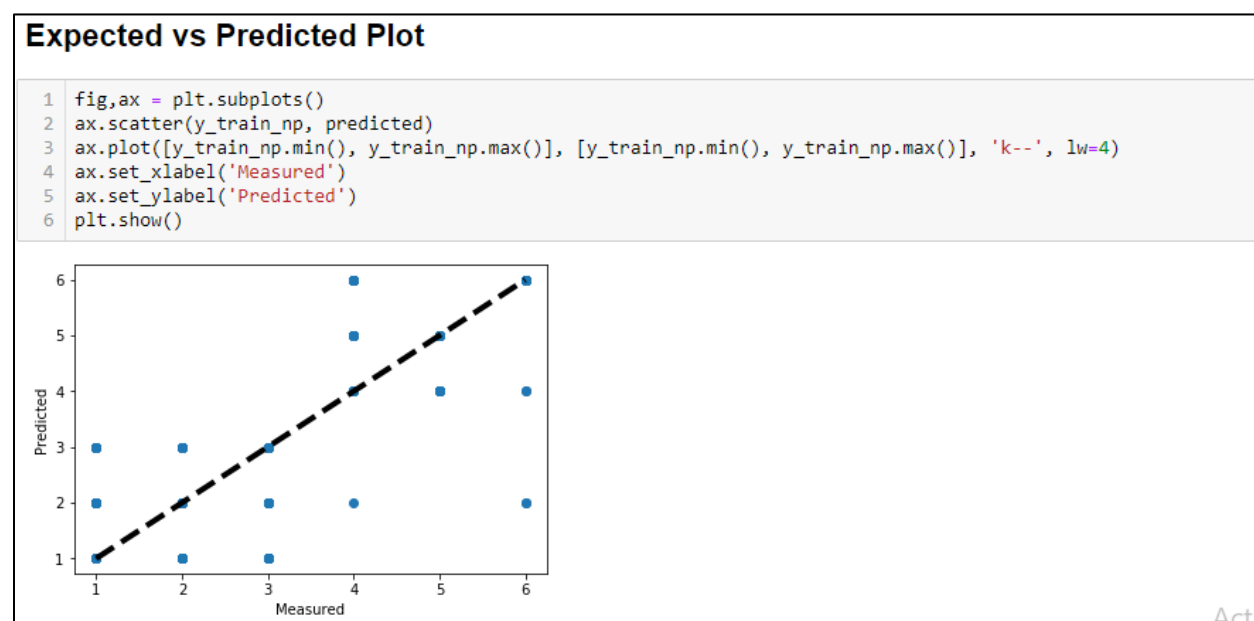


Figure 17 Scatter Plot

Neural Network

Neural networks (NN), also called artificial neural networks (ANN) are a subset of learning algorithms within the machine learning field that are loosely based on the concept of biological neural networks.

Andrey Bulezyuk, who is a German-based machine learning specialist with more than five years of experience, says that “neural networks are revolutionizing machine learning because they are capable of efficiently modelling sophisticated abstractions across an extensive range of disciplines and industries.”

Basically, an ANN comprises of the following components:

- An input layer that receives data and pass it on
- A hidden layer
- An output layers
- Weights between the layers
- A deliberate activation function for every hidden layer. In this simple neural network Python tutorial, we'll employ the Sigmoid activation function.

[How to Create a Simple Neural Network in Python \(kdnuggets.com\)](https://kdnuggets.com)

```

Epoch 9/20
7352/7352 [=====] - 9s 1ms/step - loss: 0.0192 - acc: 0.9926
Epoch 10/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0162 - acc: 0.9936
Epoch 11/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0194 - acc: 0.9923
Epoch 12/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0164 - acc: 0.9934
Epoch 13/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0170 - acc: 0.9926
Epoch 14/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0169 - acc: 0.9934
Epoch 15/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0178 - acc: 0.9931
Epoch 16/20
7352/7352 [=====] - 9s 1ms/step - loss: 0.0161 - acc: 0.9935
Epoch 17/20
7352/7352 [=====] - 9s 1ms/step - loss: 0.0124 - acc: 0.9948
Epoch 18/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0155 - acc: 0.9942
Epoch 19/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0122 - acc: 0.9951
Epoch 20/20
7352/7352 [=====] - 8s 1ms/step - loss: 0.0138 - acc: 0.9950
2947/2947 [=====] - 0s 123us/step
Accuracy: 98.53

```

- When labels were reduced to 2 only activity and no activity our neural network was able to get accuracy of nearly 99.9%
- This shows that classification of only 2 labels despite of very noisy data yields a very good accuracy.

Keras model

Keras is one of the leading high-level neural networks APIs. It is written in Python and supports multiple back-end neural network computation engines.

```

7352/7352 [=====] - 8s 1ms/step - loss: 1.0964e-06 - acc: 1.0000
Epoch 8/20
7352/7352 [=====] - 8s 1ms/step - loss: 8.0256e-07 - acc: 1.0000
Epoch 9/20
7352/7352 [=====] - 7s 1ms/step - loss: 6.2740e-07 - acc: 1.0000
Epoch 10/20
7352/7352 [=====] - 7s 998us/step - loss: 4.9859e-07 - acc: 1.0000
Epoch 11/20
7352/7352 [=====] - 8s 1ms/step - loss: 4.3401e-07 - acc: 1.0000
Epoch 12/20
7352/7352 [=====] - 8s 1ms/step - loss: 3.9036e-07 - acc: 1.0000
Epoch 13/20
7352/7352 [=====] - 8s 1ms/step - loss: 3.5395e-07 - acc: 1.0000
Epoch 14/20
7352/7352 [=====] - 8s 1ms/step - loss: 3.2262e-07 - acc: 1.0000
Epoch 15/20
7352/7352 [=====] - 8s 1ms/step - loss: 2.9227e-07 - acc: 1.0000A: 1s - loss
s - loss: 3.18
Epoch 16/20
7352/7352 [=====] - 9s 1ms/step - loss: 2.5114e-07 - acc: 1.0000
Epoch 17/20
7352/7352 [=====] - 11s 1ms/step - loss: 2.3933e-07 - acc: 1.0000
Epoch 18/20
7352/7352 [=====] - 11s 1ms/step - loss: 2.0090e-07 - acc: 1.0000
Epoch 19/20
7352/7352 [=====] - 8s 1ms/step - loss: 1.7207e-07 - acc: 1.0000
Epoch 20/20
7352/7352 [=====] - 8s 1ms/step - loss: 1.5699e-07 - acc: 1.0000
2947/2947 [=====] - 0s 86us/step
Accuracy: 99.93

```

Figure 18 Keras Model

Chapter 7

Conclusion

7.1 Results

Data analysis process is used to analyse the data in a particular way which helps to improve the quality of the system. Here 1.csv file is taken as the data analysis file which help to analyse the data. Confusion matrix is used as the accuracy test tool or matrix which helps to test the accuracy level of the data. Train and test data splitting process is used to analyse the data in a proper format which is one of the important parts of machine learning. K Fold cross validation and EDA are used to analyse the data set and find out the accuracy level of the data set. The data analysis process involves the graphical representation of the data which involves regression, classification process.

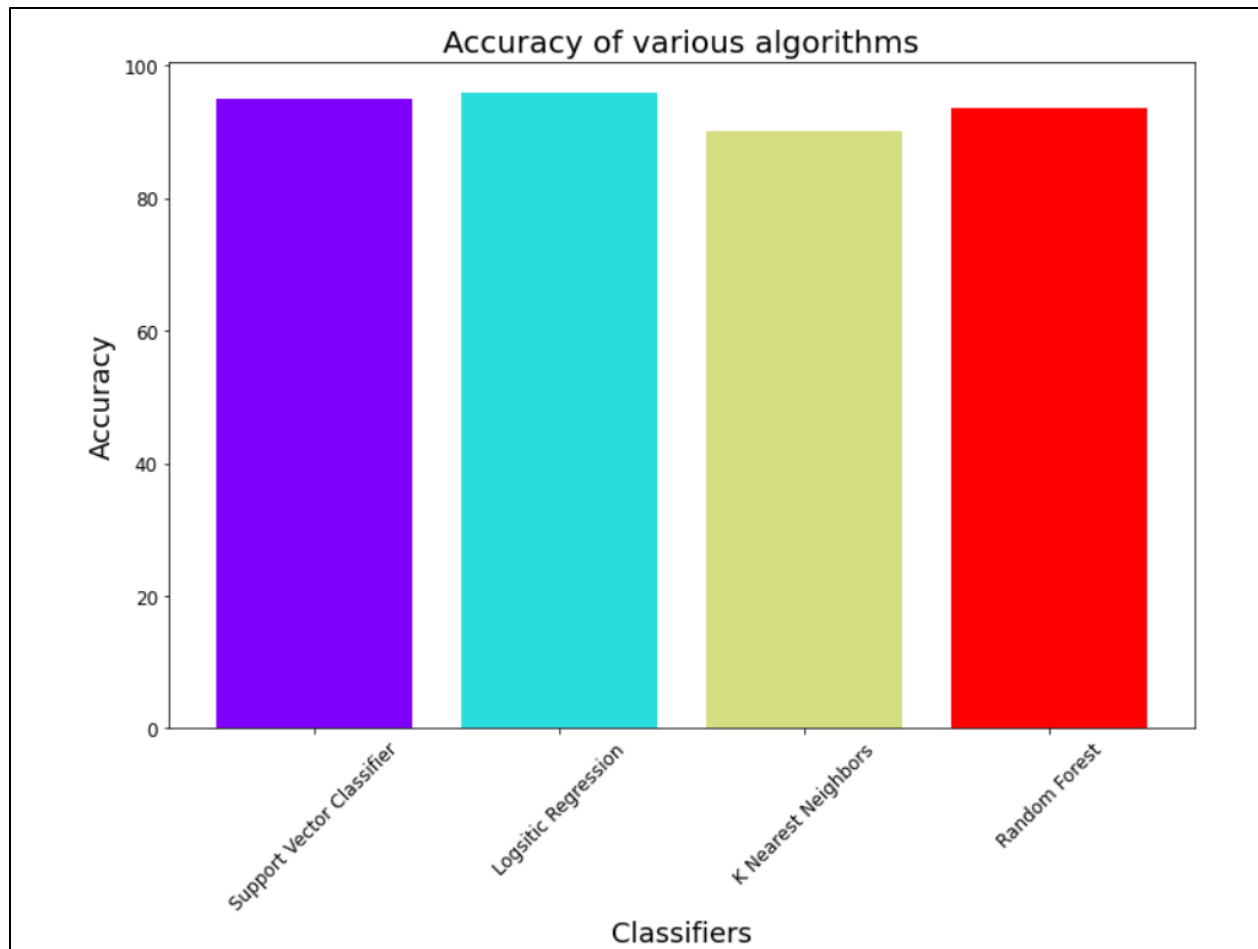


Figure 19 Accuracy of Classifiers

Logistic Regression Classifier Accuracy 96%

K- Nearest Neighbour Accuracy 89%

Support Vector Classifier Accuracy 94%

Random Forest Accuracy 91%

7.2 Research gaps

The study has been performed based analysis of standard theories and models related to the topic to meet the goals. The involvement of effective customization of design options can enhance the efficiency of the study concerning the existing process. The research has considered specific data sets to perform the modelling enabled computational system designing process. The engagement of valuable factors such as improvement of existing framework and coding can enhance the reliability of the monitoring system concerning the infrastructure development issues in this regard. Here, lack of employee support has been observed to conduct the study and optimize the quality of monitoring facilities within the existing organizational structure. The effective decision-making process is also required to be ensured in this regard to enhance the quality of outcome from the machine learning tools enabled data management and employee activity monitoring system implementation in the Alpha Bread Company concerning the other requirements.

References

- Ann, O. C. & Theng, L. B., 2014. Human activity recognition: A review. *IEEE*.
- Anon., 2017. *arrow.tudublin*. [Online] Available at: <https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1117&context=scschcomdis> [Accessed 1 January 2021].
- ChhaviDhimana, Kumar, D. & Vishwakarmab, 2019. A review of state-of-the-art techniques for abnormal human activity recognition. *Engineering Applications of Artificial intelligence*, p. 45.
- Croarkin, C. T. P. F. J. e. a., 2006. *Nist/sematech e-handbook of statistical methods*. [Online] Available at: <http://www.itl.nist.gov/div898/handbook> [Accessed 5 January 2021].
- De-La-Hoz-Franco, E., Ariza-Colpas, P., Quero, J. M. & Espinilla, M., 2018. Sensor-Based Datasets for Human Activity Recognition – A Systematic Review of Literature. *IEEE*.
- Grubbs, F. E., 2012. Procedures for detecting outlying observations in samples. *Procedures for Detecting Outlying Observations in Samples*, 11(1), pp. 1-21.
- Jozsef Suto, S. O. C. L. & I. O., 2018. Comparison of offline and real-time human activity recognition results using machine learning techniques. *Advances in Bio Inspired Intelligent systems*.
- MaryamZiaeeefard & RobertBergevin, 2015. Pattern Recognition -Science Direct. *Semantic human activity recognition: A literature review*, pp. 2329-2345.
- Michael D. Blechner, M., 2005. *Behavior of Various Machine Learning Models in the Face of Noisy Data*, s.l.: Harvard-MIT Division of Health Sciences and Technology.
- Netzahualcoyotl Hernandez, J. L. J. F. I. M. & B. A., 2020. Literature Review on Transfer Learning for Human Activity Recognition Using Mobile and Wearable Devices with Environmental Technology. *Springer Link*.
- Ranasinghe, S., Machot, F. A. & Mayr, H. C., 2016. A review on applications of activity recognition systems with regard to performance and evaluation. *International Journal of Distributed Sensor Networks*.

Shugang Zhang, I. Z. W. , J. N. L. H. , S. W. , a. Z. L., 2017. A Review on Human Activity Recognition Using Vision-Based Method. *Computer Vision in Healthcare Applications*.

Sousa, W. et al., 2017. A Comparative Analysis of the Impact of Features on Human Activity Recognition with Smartphone Sensors. *A Comparative Analysis of the Impact of Features on Human Activity Recognition with Smartphone Sensors*, pp. 397-404.

Vellampalli, H., 2017. *Arrow@TU Dublin*. [Online] Available at:

<https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1117&context=scschcomdis>

[Accessed 1 January 2021].

Vellampalli, H., 2017. Physical Human Activity Recognition Using Machine Learning. *Arrow TU DUBLIN*, 10 November, pp. 28-46.

Vrigkas, M., Nikou, C. & Kakadiaris, I. A., 2015. A Review of Human Activity Recognition Methods. *Frontiers in Robotics and AI*.

Zhu, X. & W. X., 2004. Class noise vs. attribute noise: A quantitative study. *Artificial Intelligence Review*, 22(3), pp. 177-210.

Sondhi, P. (2009). Feature construction methods: a survey. *sifaka. cs. uiuc. edu*, 69, 70–71.

Starner, T., Weaver, J., & Pentland, A. (1997). A wearable computer-based american sign language recogniser. *Personal Technologies*, 1 (4), 241–250.

Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., & Tröster, G. (2008). Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, 7 (2).

Sucerquia, A., López, J. D., & Vargas-Bonilla, J. F. (2017). Sisfall: a fall and movement dataset. *Sensors*, 17 (1), 198.

Sung, M., DeVaul, R., Jimenez, S., Gips, J., & Pentland, A. (2004). Shiver motion and core body temperature classification for wearable soldier health monitoring systems. In *Wearable computers, 2004. iswc 2004. eighth international symposium on* (Vol. 1, pp. 192–193).

Tapia, E. M., Intille, S. S., & Larson, K. (2004). Activity recognition in the home using simple and ubiquitous sensors. In *Pervasive* (Vol. 4, pp. 158–175).

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, 6 (1), 35-39. Retrieved from <http://dx.doi.org/10.1177/875647939000600106> doi: 10.1177/875647939000600106

Tessendorf, B., Bulling, A., Roggen, D., Stiefmeier, T., Feilner, M., Derleth, P., & Tröster, G. (2011). Recognition of hearing needs from body and eye movements to improve hearing instruments. *Pervasive Computing*, 314–331.

Turaga, P., Chellappa, R., Subrahmanian, V. S., & Udrea, O. (2008, November). Machine recognition of human activities: A survey. *IEEE Trans. Cir. and Sys. for Video Technol.*, 18 (11), 1473–1488. Retrieved from <http://dx.doi.org/10.1109/TCSVT.2008.2005594> doi: 10.1109/TCSVT.2008.2005594

Wan, D. (1999). Magic medicine cabinet: A situated portal for consumer healthcare. In *International symposium on handheld and ubiquitous computing* (pp. 352–355).

Liberty, E., Lang, K., & Shmakov, K. (2016). Stratified sampling meets machine learning. In *International conference on machine learning* (pp. 2320–2329).

Vaníčková, R., 2017. Application of PRINCE2 project management methodology. *StudiaCommercialiaBratislavensia*, 10(38), pp.227-238.

Wang, C., Cheung, A. and Bodik, R., 2017, June. Synthesizing highly expressive SQL queries from input-output examples. In *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation* (pp. 452-466).

Wang, S., Wu, Y.C., Xia, M., Wang, R. and Poor, H.V., 2020. Machine intelligence at the edge with learning centric power allocation. *IEEE Transactions on Wireless Communications*, 19(11), pp.7293-7308.

Widodo, C.E., Adi, K. and Gernowo, R., 2020, April. Medical image processing using python and open cv. In *Journal of Physics: Conference Series* (Vol. 1524, No. 1, p. 012003). IOP Publishing.

Wuest, T., Weimer, D., Irgens, C. and Thoben, K.D., 2016. Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1), pp.23-45.

Zhang, L., Tan, J., Han, D. and Zhu, H., 2017. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today*, 22(11), pp.1680-1685.

Zhang, Y., Genkin, D., Katz, J., Papadopoulos, D. and Papamanthou, C., 2017, May. vSQL: Verifying arbitrary SQL queries over dynamic outsourced databases. In *2017 IEEE Symposium on Security and Privacy (SP)* (pp. 863-880). IEEE.

Zhao, Y., Yang, R., Chevalier, G., Xu, X. and Zhang, Z., 2018. Deep residual bidir-LSTM for human activity recognition using wearable sensors. *Mathematical Problems in Engineering*, 2018, pp. 632.