**IBM PROJECT REPORT**

**Project Name**: Detecting Parkinson’s Disease using Machine Learning.

**Team ID**: PNT2022TMID04005

**Team**: Maneesh Vijay V I (Team Lead)

Hariharan K

Sivakumar M S

Kishore Kumar M

1. **INTRODUCTION**
   1. Project Overview
   2. Purpose
2. **IDEATHON AND PROPOSED SOLUTION**
   1. Empathy Map
   2. Literature Survey
   3. Problem Statement Definition
   4. Brainstorming
3. **PROJECT DESIGN PHASE-I**
   1. Problem Solution Fit
   2. Proposed Solution
   3. Solution Architecture
4. **PROJECT DESIGN PHASE-II**
   1. Customer Journey
   2. Functional Requirement
   3. Data Flow Diagrams and User Stories
   4. Technology Architecture
5. **PROJECT PLANNING PHASE** 
   1. Milestones and Activity List
   2. Sprint Delivery
6. **CONCLUSION**

**1.INTRODUCTION**

**Detecting Parkinson’s Disease using Machine Learning.**

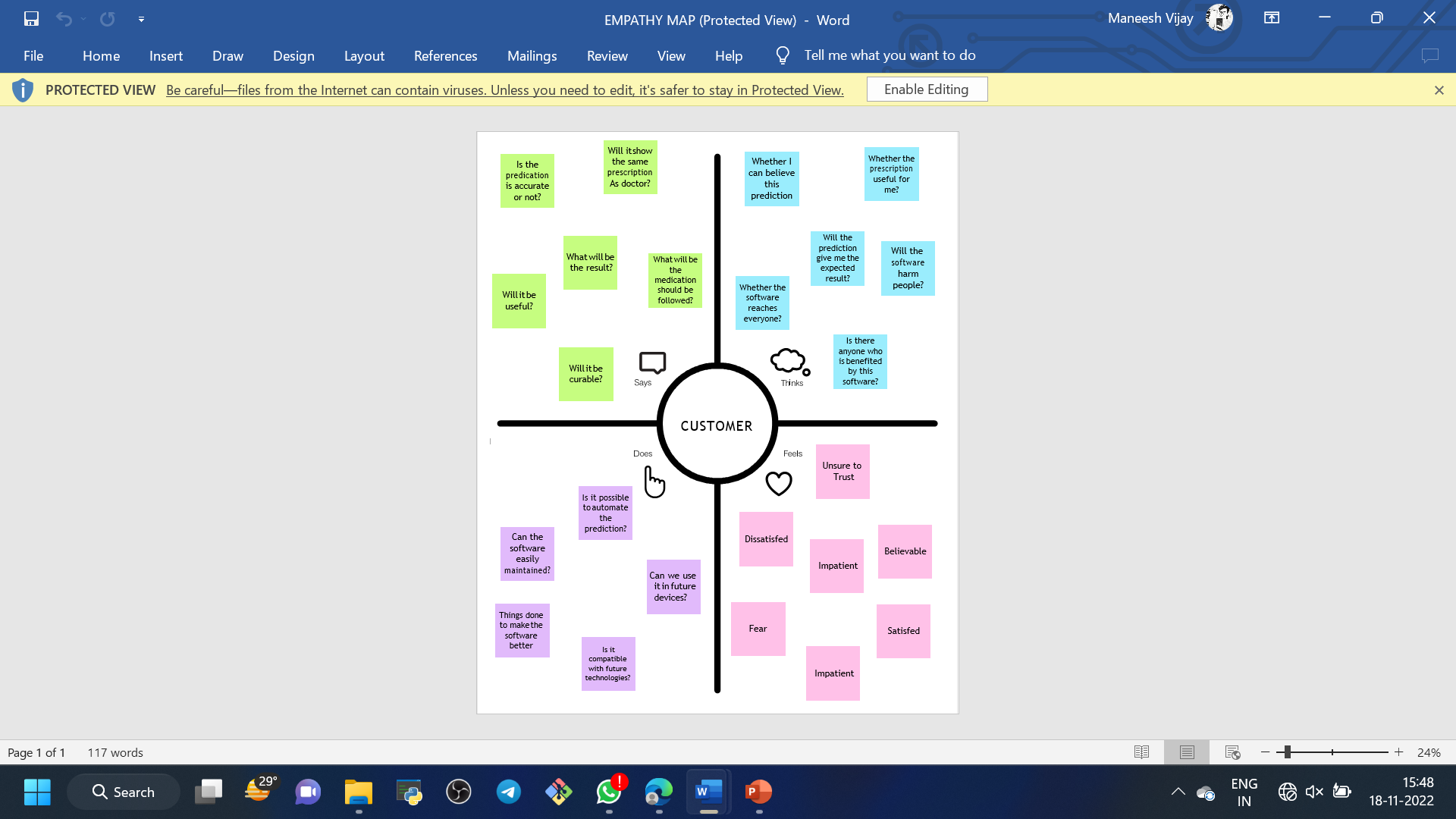
Parkinson Disease is a brain neurological disorder. It leads to shaking of the body, hands and provides stiffness to the body. No proper cure or treatment is available yet at the advanced stage. Treatment is possible only when done at the early or onset of the disease. These will not only reduce the cost of the disease but will also possibly save a life. Most methods available can detect Parkinson in an advanced stage; which means loss of approx.. 60% dopamine in basal ganglia and is responsible for controlling the movement of the body with a small amount of dopamine. More than 145,000 people have been found alone suffering in the U.K and in India, almost one million population suffers from this disease and it’s spreading fast in the entire world.

**PROJECT FLOW:**

* User interacts with the UI (User Interface) to upload the image as input
* The uploaded image is analyzed by the model which is integrated
* Once the model analyzes the uploaded image, the prediction is showcased on the UI and OpenCV window
* To accomplish this, we have to complete all the activities and tasks listed below
* Data Collection
* Image Preprocessing.
* Importing the required libraries
* Loading Train data and Test data
* Quantifying images
* Label Encoding
* Model Building
* Application Building

**2. IDEATHON AND PROPOSED SOLUTION**

**EMPATHY MAP**



**LITERATURE SURVEY**

**Survey 1:**

**Daiga Heisters (2011):**

**Parkinson’s: symptoms, treatments and research:**

Daiga Heisters et al. proposed the paper titled “Parkinson’s: symptoms, treatments and research”. This paper initially says that Current treatments can help to ease the symptoms but none can repair the damage in the brain or slow the progress of the condition; now, Parkinson’s UK researchers are working to develop new treatments that can and finally worked together to 12 build on existing discoveries and explore these innovative areas of research, it is hoped that a cure for Parkinson’s will be found. Parkinson’s UK offers support for everyone affected,, including people with the condition, their family, friends and careers, researchers and professionals working in this area.

# Survey 2:

**Afzal Hussain Shahid and Maheshwari Prasad Singh (2016):**

**A deep learning approach for prediction of Parkinson’s disease progression:**

Afzal Hussain Shahid and Maheshwari Prasad Singh et al. proposed the paper titled “A deep learning approach for prediction of Parkinson’s disease progression” . This paper proposed a deep neural network (DNN) model using the reduced input feature space of Parkinson’s telemonitoring dataset to predict Parkinson’s disease (PD) progression and also proposed a 11 PCA based DNN model for the prediction of Motor-UPDRS and Total-UPDRS in Parkinson's Disease progression. The DNN model was evaluated on a real-world PD dataset taken from UCI. Being a DNN model, the performance of the proposed model may improve with the addition of more data points in the datasets.

# Survey 3:

**T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi (2018):**

**Parkinson’s Disease Diagnosis Using Machine Learning and Voice:**

T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi et al., proposed the paper titled “Parkinson’s Disease Diagnosis Using Machine Learning and Voice” .It is that it explores the effectiveness of using supervised classification algorithms, such as deep neural networks, to accurately diagnose individuals with the disease. Historically, PD has been difficult to quantify and doctors have tended to focus on some symptoms while ignoring others, relying primarily on subjective rating scales. The analysis of this paper provides a comparison of the effectiveness of various machine learning classifiers in disease diagnosis with noisy and high dimensional data. Their peak accuracy of 85% provided by the machine learning models exceeds the average clinical diagnosis accuracy of non-experts (73.8%) and average accuracy of movement disorder specialists (79.6% without follow-up, 83.9% after follow-up) with pathological post-mortem examination as ground truth.

# Survey 4:

**Mohamad Alissa (2018):**

**Parkinson’s Disease Diagnosis Using Deep Learning:**

Mohamad Alissa et al. Proposed the paper titled “Parkinson’s Disease Diagnosis Using Deep Learning” . This project mainly aims to automate the PD diagnosis process using deep learning, Recursive Neural Networks (RNN) and Convolutional Neural Networks (CNN), to differentiate between healthy and PD patients. Besides that, since different datasets may capture different aspects of this disease, this project aims to explore which PD test is more effective in the discrimination process by analysing different imaging and movement datasets (notably cube and spiral pentagon datasets). In general, the main aim of this paper is to automate the PD diagnosis process in order to discover this disease as early as possible. If we discover this disease earlier, then the treatments are more likely to improve the quality of life of the patients and their families.

There are some limitations to this paper namely:

* They used the validation set only to investigate the model performance during the training and this reduced the number of samples in the training set.
* RNN training is too slow and this is not flexible in practice work.
* Disconnecting and resource exhaustion: working with cloud services like Google Collaboratory causes many problems like disconnecting suddenly. And because it is shareable service by the world zones, this leads to resource exhaustion error many times.

# Survey 5:

**T. Swapna, Y. Sravani Devi (2019):**

**Performance Analysis of Classification algorithms on Parkinson’s Dataset with Voice Attributes:**

T. Swapna, Y. Sravani Devi et al. proposed a paper and titled “Performance Analysis of Classification algorithms on Parkinson’s Dataset with Voice Attributes” .This paper deals with the application of seven classification algorithms on the acquired data set and then drawing out a comparison of the results to one another and also predicting the outcome whether the person is healthy or Parkinson disease effected from the given data. The results of the selected algorithms namely Naïve Bayes, Random Forest, Neural Networks, Decision Trees, AdaBoost, SVM, KNN, LGBM were compared and tabulated. According to the outputs derived with the help of python, implementing Scikit Libraries. Final accuracy was calculated using these parameters. Random Forest algorithm gives with optimum accuracy of 78.56% which is closely followed by Decision Tree Algorithm with the optimal accuracy of 77.63%. Following the Decision Tree Algorithm is the MLP Classifier with an optimal accuracy of 76.72%, and the Naïve Bayes Algorithm which has the optimal accuracy of 70.82% and lastly Light Gradient Boosting Model has the optimal accuracy of 90% Finally, this algorithm can help in classifying whether a person get affected with Parkinson’s disease or not.

# Survey 6:

**Anila M and Dr G Pradeepini (2020):**

**Diagnosis of Parkinson’s disease using Artificial Neural network:**

Anila M and Dr G Pradeepini et al. proposed the paper titled “Diagnosis of Parkinson’s disease using Artificial Neural network”. The main objective of this paper is that the detection of the disease is performed by using the voice analysis of the people affected with Parkinson's disease. For this purpose, various machine learning techniques like ANN, Random Forest, KNN, SVM, XG Boost are used to classify the best model, error rates are calculated, and the performance metrics are evaluated for all the models used. The main drawback of this paper is that it is limited to ANN with only two hidden layers. And this type of neural networks with two hidden layers are sufficient and efficient for simple datasets. They used only one technique for feature selection which reduces the number of features.

# Survey 7:

**Arvind Kumar Tiwari (2020):**

**Machine Learning-based Approaches for Prediction of Parkinson’s Disease:**

Arvind Kumar Tiwari et al. proposed the paper titled “Machine Learning-based Approaches for Prediction of Parkinson’s Disease” . In this paper, minimum redundancy maximum relevance feature selection algorithms were used to select the most important feature among all the features to predict Parkinson diseases. Here, it was observed that the random forest with 20 10 number of features selected by minimum redundancy maximum relevance feature selection algorithms provide the overall accuracy 90.3%, precision 90.2%, Mathews correlation coefficient values of

0.73 and ROC values 0.96 which is better in comparison to all other machine learning based approaches such as bagging, boosting, random forest, rotation forest, random subspace, support vector machine, multilayer perceptron, and decision tree based methods.

# Survey 8:

**Siva Sankara Reddy Donthi Reddy and Udaya Kumar Ramanadham (2020): Prediction of Parkinson’s Disease at Early Stage using Big Data Analytics:**

Siva Sankara Reddy Donthi Reddy and Udaya Kumar Ramanadham et al. proposed the paper “Prediction of Parkinson’s Disease at Early Stage using Big Data Analytics” . This paper describes mainly various Big Data Analytical techniques that may be used in diagnosing of right disease in the right time. The main intention is to verify the accuracy of prediction algorithms. Their future study aims to propose an efficient method to diagnose this type of neurological disorder by some symptoms at the early stage with better accuracy using different Big Data Analytical techniques like Hadoop, Hive, R Programming, MapReduce, PIG, Zookeeper, HBase, Cassandra, Mahout etc…

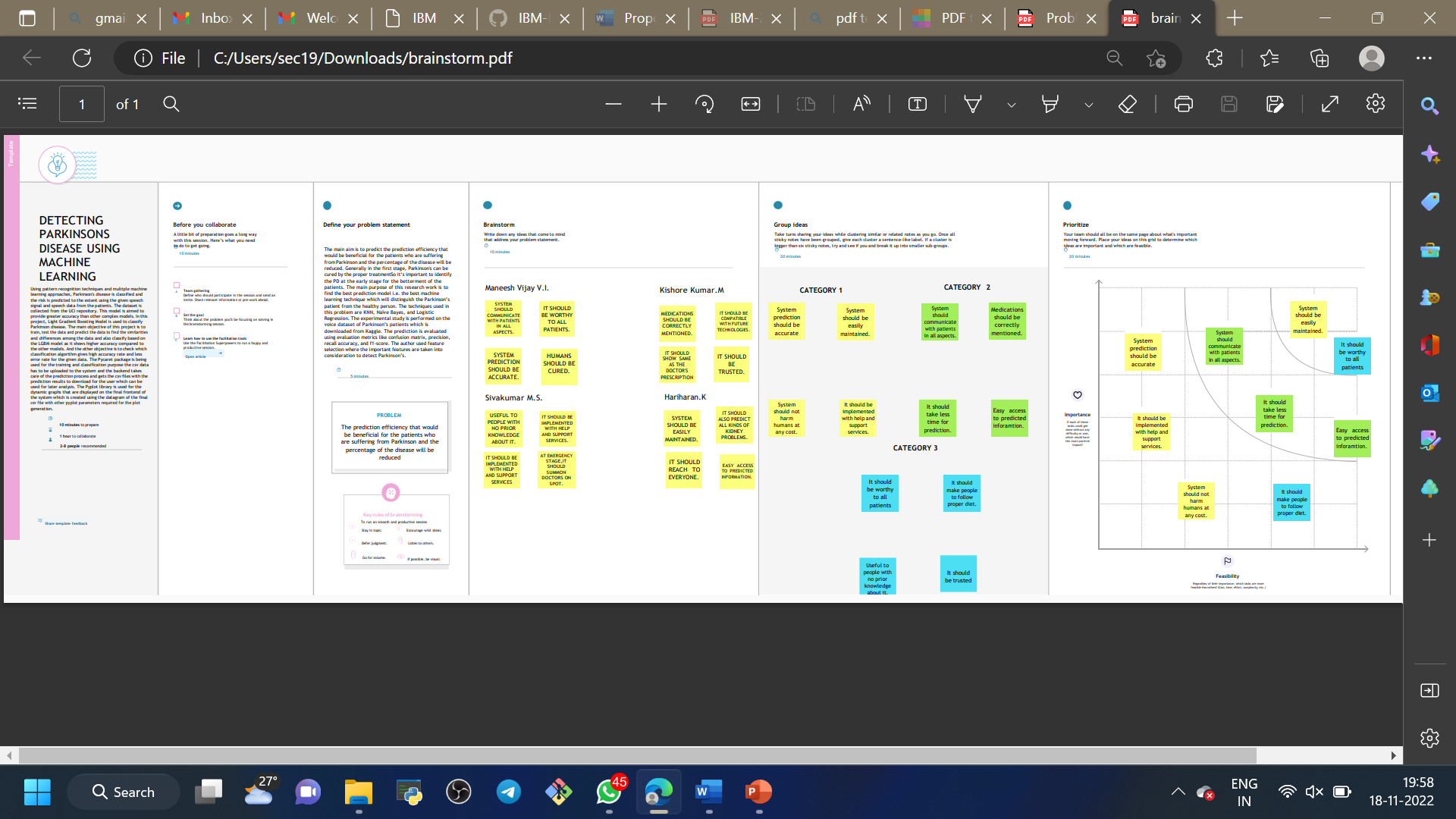
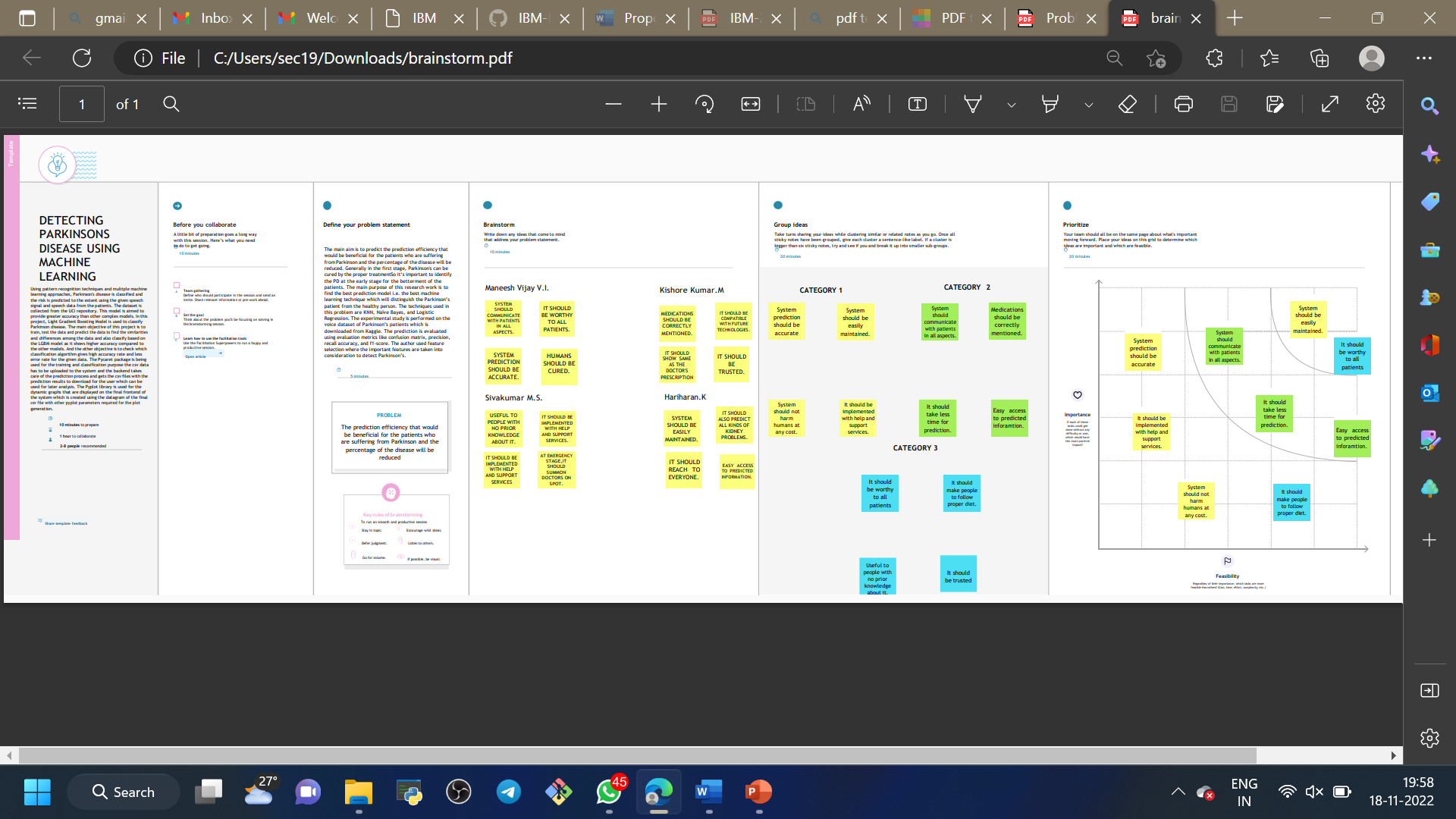
**REFERENCES**

1. Heisters. D, “Parkinson’s: symptoms, treatments and research”. British Journal of Nursing, 20(9), 548–554. doi:10.12968/bjon.2011.20.9.548, 2011.
2. Arvind Kumar Tiwari, “Machine Learning based Approaches for Prediction of Parkinson’s Disease” Machine Learning and Applications: An International Journal (MLAU) vol. 3, June 2016.
3. T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins and R. H. Ghomi, "Parkinson’s Disease Diagnosis Using Machine Learning and Voice," IEEE Signal Processing in Medicine and Biology Symposium (SPMB), pp.1-7, doi: 10.1109/SPMB.2018.8615607, 2018.
4. Mohamad Alissa,” Parkinson’s Disease Diagnosis Using Deep Learning”, August 2018.
5. T. Swapna, Y. Sravani Devi, “Performance Analysis of Classification algorithms on Parkinson’s Dataset with Voice Attributes”. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 2 pp. 452-458, 2019.
6. Anila M Department of CS1, Dr G Pradeepini Department of CSE, “DIAGNOSIS OF PARKINSON’S DISEASE USING ARTIFICIAL NEURAL NETWORK”, JCR, 7(19): 7260-7269, 2020.
7. Shahid, A.H., Singh, M.P. A deep learning approach for prediction of Parkinson’s disease progression, https://doi.org/10.1007/s13534-020-00156-7, Biomed. Eng. Lett. 10, 227– 239, 2020.
8. Siva Sankara Reddy Donthi Reddy and Udaya Kumar Ramanadham “Prediction of Parkinson’s Disease at Early Stage using Big Data Analytics”ISSN: 2249 – 8958, Volume- 9 Issue-4, April 2020.

**PROBLEM STATEMENT DEFINITION**

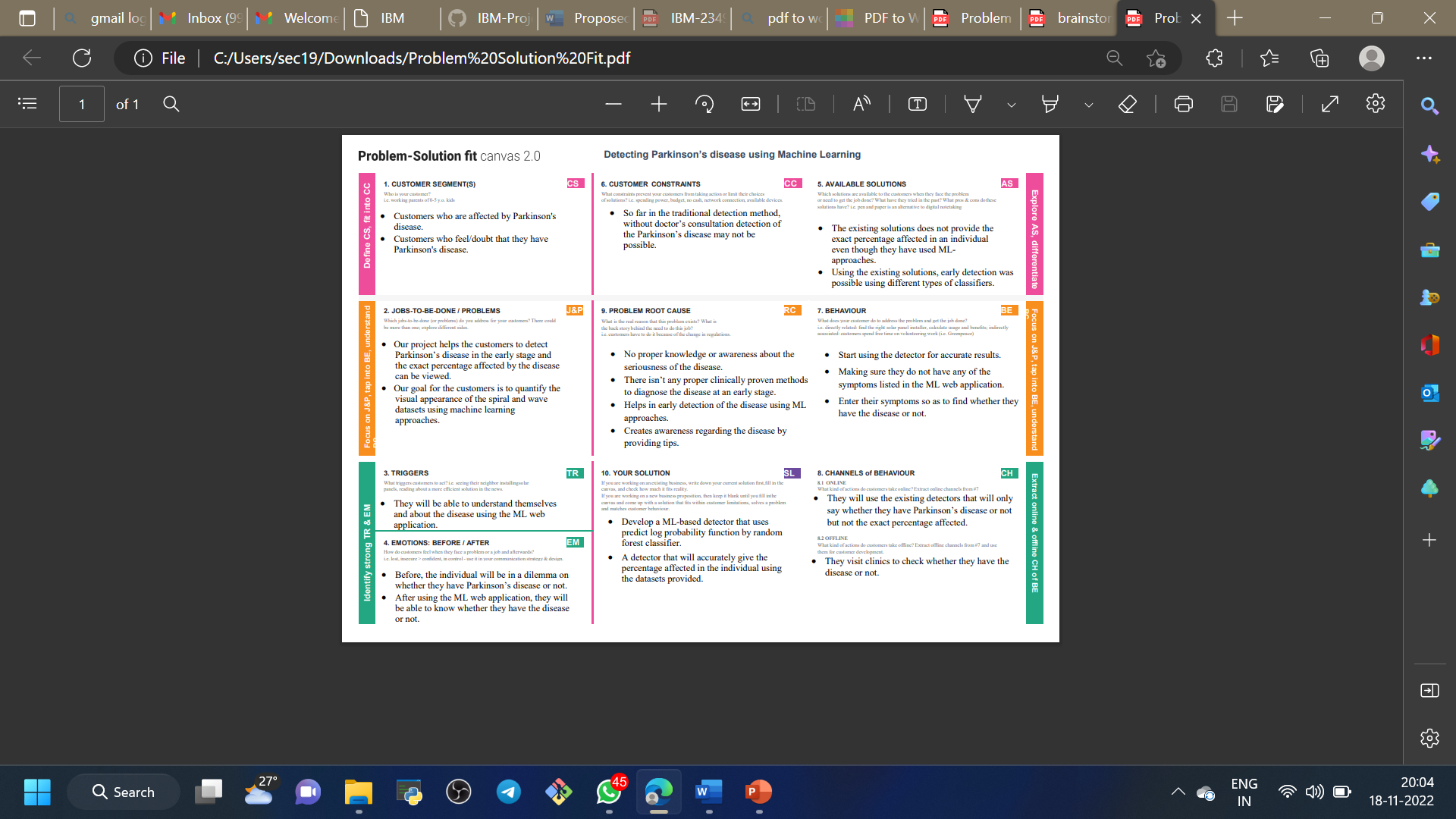
The main aim is to predict the prediction efficiency that would be beneficial for the patients who are suffering from Parkinson and the percentage of the disease will be reduced. Generally in the first stage, Parkinson's can be cured by the proper treatmentSo it‘s important to identify the PD at the early stage for the betterment of the patients. The main purpose of this research work is to find the best prediction model i.e. the best machine learning technique which will distinguish the Parkinson’s patient from the healthy person. The techniques used in this problem are KNN, Naïve Bayes, and Logistic Regression. The experimental study is performed on the voice dataset of Parkinson’s patients which is downloaded from Kaggle. The prediction is evaluated using evaluation metrics like confusion matrix, precision, recall accuracy, and f1-score. The author used feature selection where the important features are taken into consideration to detect Parkinson’s disease.

**BRAINSTORMING**

**3. PROJECT DESIGN PHASE-I**

**PROBLEM SOLUTION FIT**



**PROPOSED SOLUTION**

**Project Design Phase-I**

**Proposed Solution Template**

|  |  |
| --- | --- |
| Date | 03 November 2022 |
| Team ID | PNT2022TMID04005 |
| Project Name | Detecting Parkinsons Disease Using Machine Learning |
| Maximum Marks | 2 Marks |

**Proposed Solution Template:**

Project team shall fill the following information in proposed solution template.

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Description** |
|  | Problem Statement (Problem to be solved) | Parkinson’s disease disorder is a brain disorder that causes unintended or uncontrollable movements, such as shaking, stiffness, and difficulty with balance and coordination. Symptoms usually begin gradually and worsen over time. As the disease progresses, people may have difficulty walking and talking. |
|  | Idea / Solution description | Studies investigates signals from sustained phonation and text dependent speech modalities for parkinson's disease screening. Phonation corresponds to the vowel voicing task and speech to the pronounciation of a short sentence, signal will be recorded through channel simultaneously through mobile phone or microphone. Parkinson disease affect vocal chord so the motion of speech is detected and evaluated. |
|  | Novelty / Uniqueness | Testing 25 non impulsive patients with Parkinson's disease (PD) and 27 PD patients with impulsive compulsive behaviours (ICBs). Both patient groups were examined "on" and "off" dopaminergic medication in a counterbalanced order and their behaviour was compared with 24 healthy controls. We found that PD patients with ICBs were significantly more prone to choose novel options than either non impulsive PD patients or controls, regardless of medication status. Our findings suggest that attraction to novelty is a personality trait in all PD patients with ICBs which is independent of medication status. |
|  | Social Impact / Customer Satisfaction | Since it is based on the voice based detection it is very convenient to use. As it helps the people to detect the Parkinson’s disease in early stage, the loss of life is prevented. It detect without cost and helps to avoid travelling and time. |
|  | Business Model (Revenue Model) | A free platform with useful feature. Any adult and young people can use it and suggest it to others to increase the value |
|  | Scalability of the Solution | Additional features can be added anytime anywhere. Any number of users can access it all at once. |

**SOLUTION ARCHITECTURE**

**Project Design Phase-I**

**Solution Architecture**

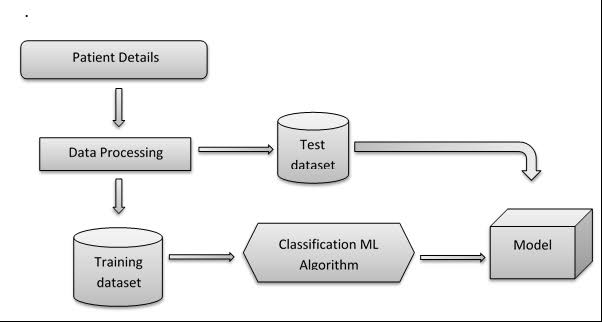
|  |  |
| --- | --- |
| Date | 03 November 2022 |
| Team ID | PNT2022TMID04005 |
| Project Name | Detecting Parkinsons Disease using Machine Learning |
| Maximum Marks | 4 Marks |

**Solution Architecture:**

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

* Find the best tech solution to solve existing business problems.
* Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.
* Define features, development phases, and solution requirements.
* Provide specifications according to which the solution is defined, managed, and delivered.

**Example - Solution Architecture Diagram:**

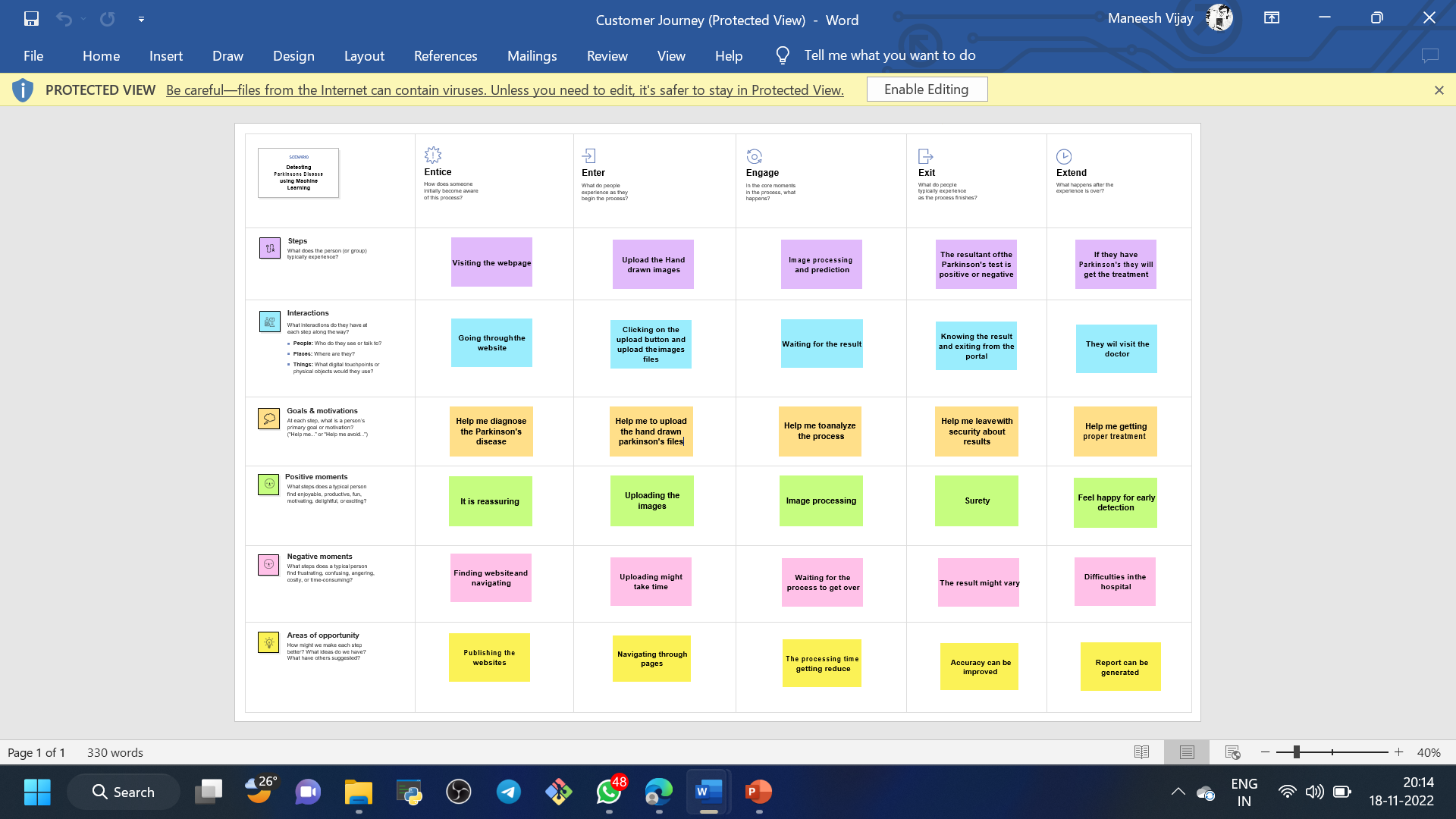


*Figure 1: Architecture and data flow of the voice patient diary sample application*

**Reference:** [**https://aws.amazon.com/blogs/industries/voice-applications-in-clinical-research-powered-by-ai-on-aws-part-1-architecture-and-design-considerations/**](https://aws.amazon.com/blogs/industries/voice-applications-in-clinical-research-powered-by-ai-on-aws-part-1-architecture-and-design-considerations/)

**4. PROJECT DESIGN PHASE-II**

**CUSTOMER JOURNEY**



# 

**FUNCTIONAL REQUIREMENTS/SOLUTION REQUIREMENTS**

# Project Design Phase-II

**Solution Requirements (Functional & Non-functional)**

|  |  |
| --- | --- |
| Date | 03 November 2022 |
| Team ID | PNT2022TMID04005 |
| Project Name | Detecting Parkinson’s disease using machine learning |
| Maximum Marks | 4 Marks |

## 

## Functional Requirements:

Following are the functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Functional Requirement (Epic)** | **Sub Requirement (Story / Sub-Task)** |
| FR-1 | environment independent | It often a calculation, data manipulation, business process, user interaction, or the other specific  functionality which defines what function a system is probably going to perform. |
| FR-2 | Testing | Applying the algorithms on the test data |
| FR-3 | Confirmation | Display the result with the description of having Parkinson’s or not. |

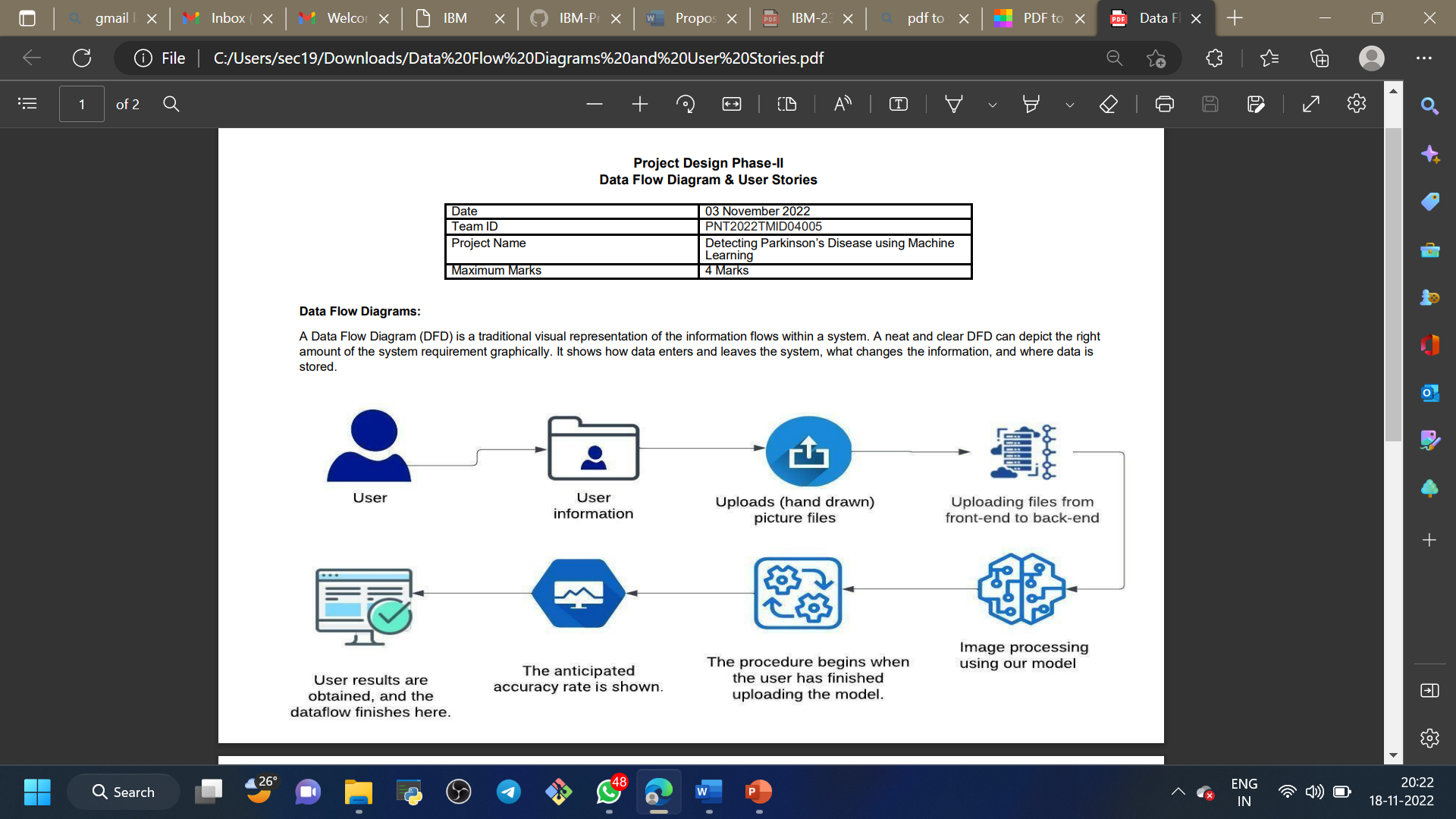
## 

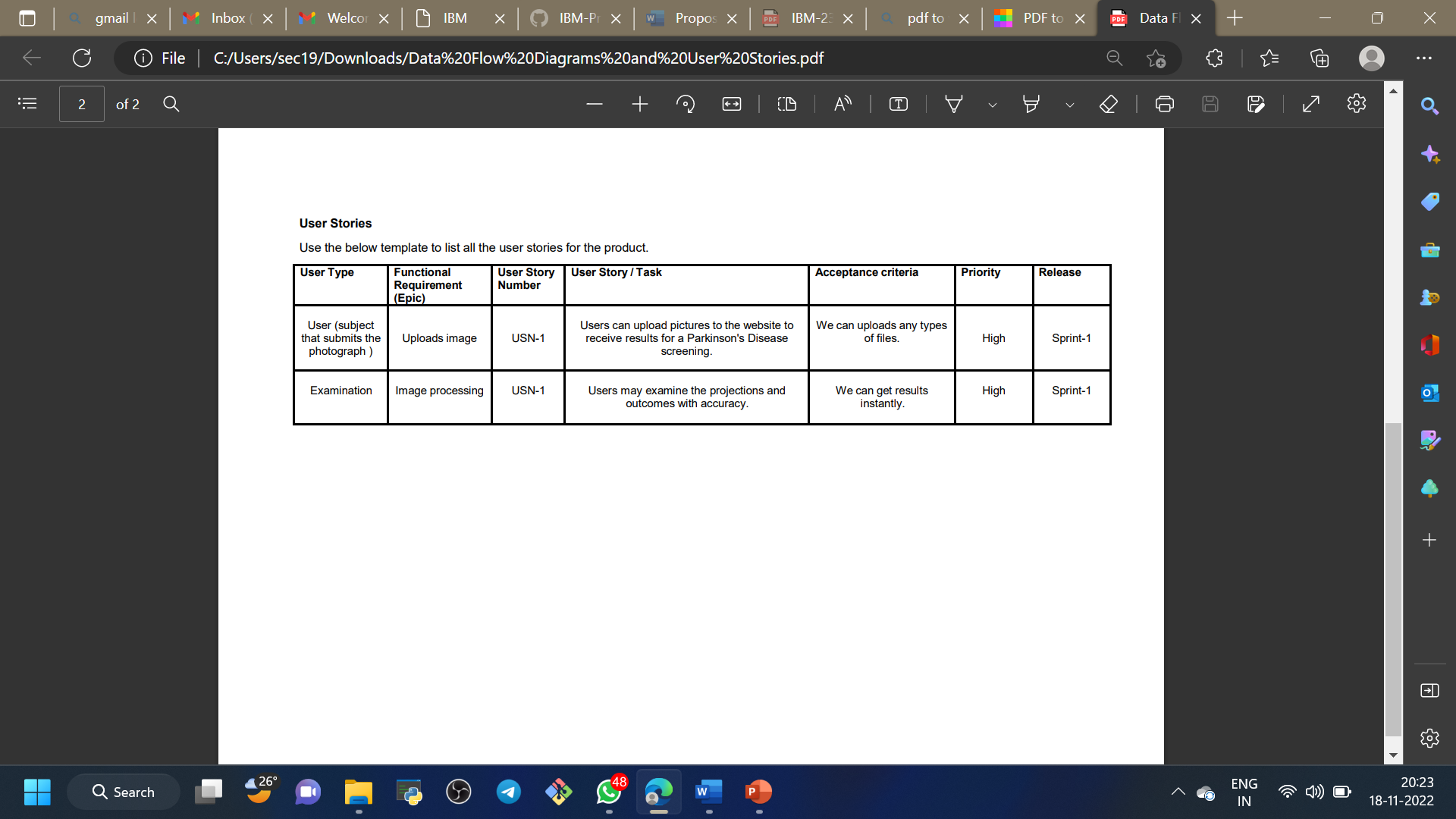
## Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Non-Functional Requirement** | **Description** |
| NFR-1 | **Usability** | The nervous system conditions (neurologist) will diagnose Parkinson's disease based on your medical history, a review of your signs and  symptoms, and a neurological and physical examination. |
| NFR-2 | **Security** | Parkinson symptoms make falls more likely. Safety improvements around the house can help. But if you start having frequent falls, talk to  your doctor. He or she may recommend physical therapy. |
| NFR-3 | **Reliability** | Low diagnostic accuracy is particularly relevant in the early stages of disease and presumably in older patients. |
| NFR-4 | **Performance** | This enables easy and early detection of disease which can significantly improve symptoms and  quality of life. |
| NFR-5 | **Availability** | The system uses the drawing of a person and analyses the pattern. |
| NFR-6 | **Scalability** | It can be implemented using any we framework and  we framework and we can be made available to everyone in need. |

**DATA FLOW DIAGRAMS AND USER STORIES**

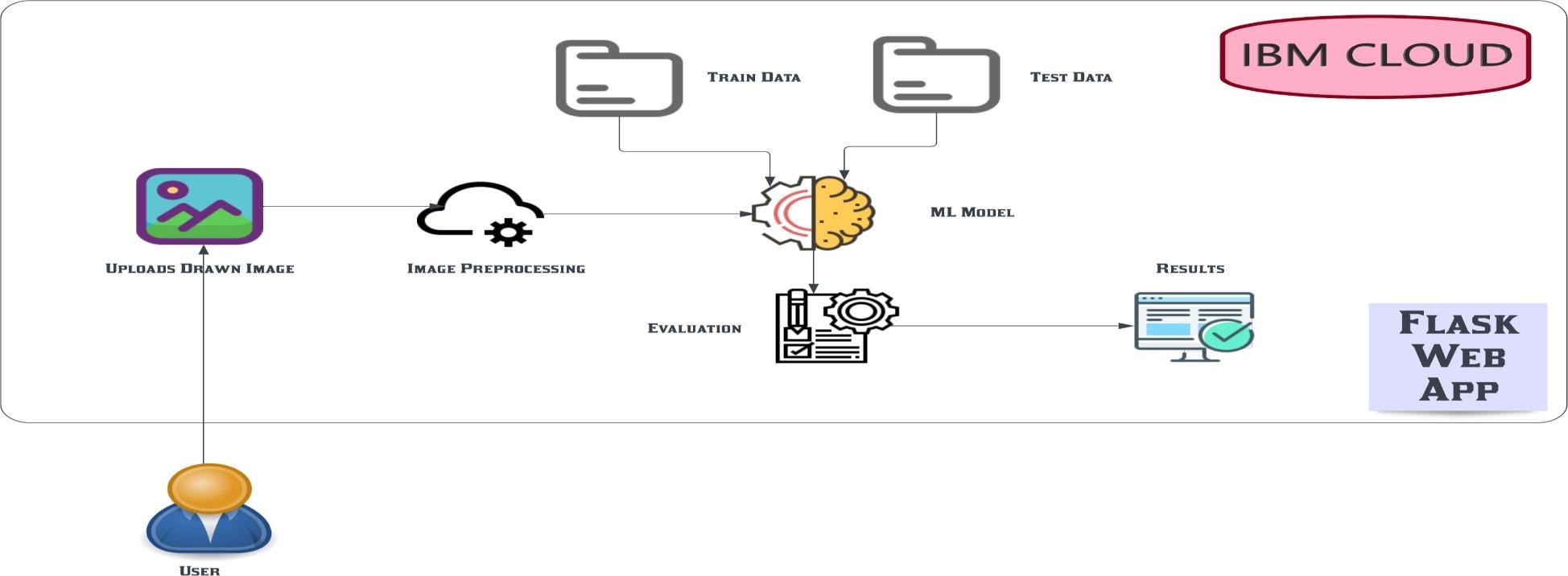


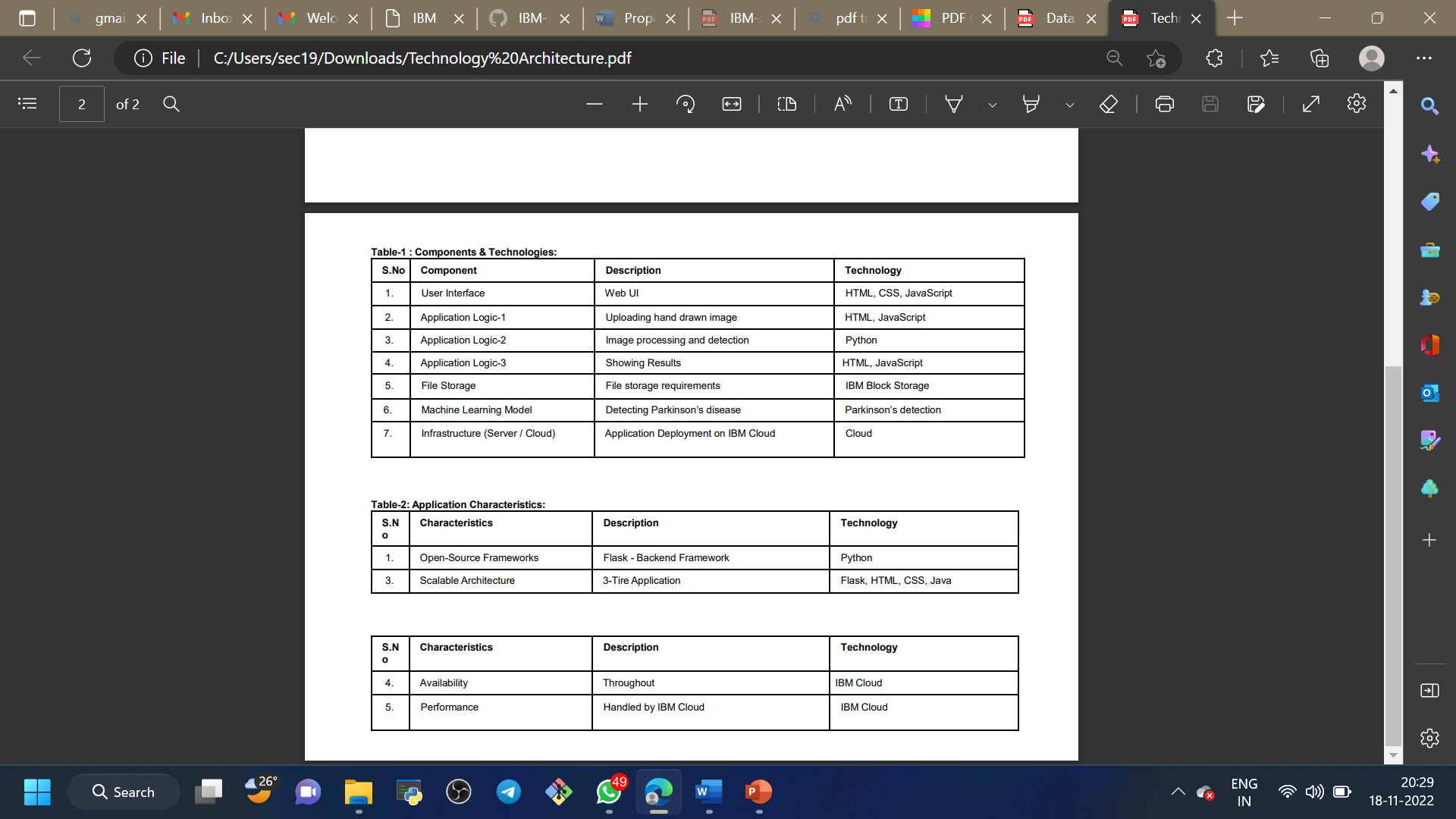
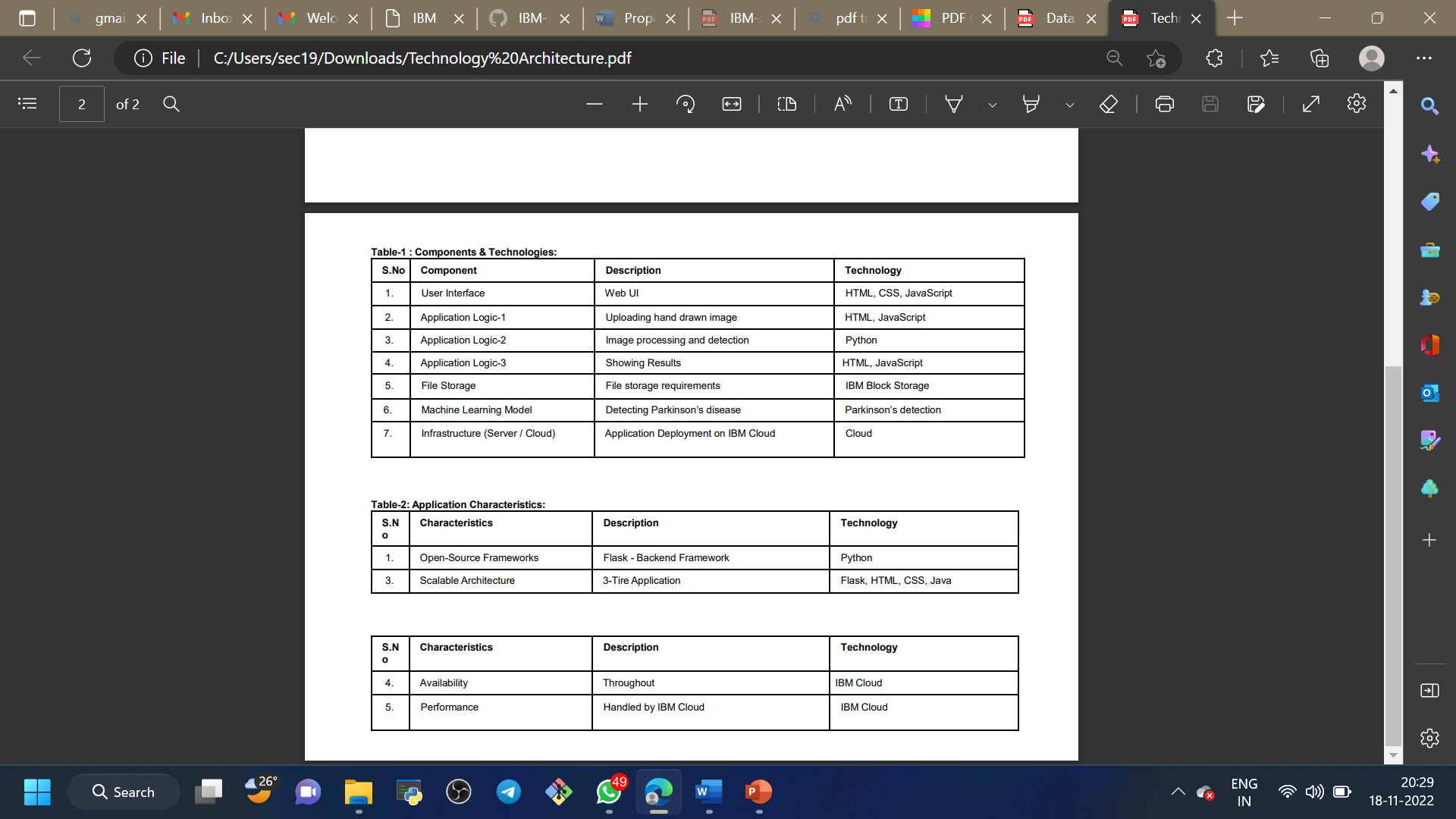


**TECHNOLOGY ARCHITECTURE**

Project Design Phase-II Technology Stack (Architecture & Stack)

|  |  |
| --- | --- |
| Date | 03 November 2022 |
| Team ID | PNT2022TMID04005 |
| Project Name | Detecting Parkinson’s Disease using Machine Learning |
| Maximum Marks | 4 Marks |



**5.PROJECT PLANNING PHASE**

**MILESTONES AND ACTIVITY LIST**

Project Planning Phase-Milestones and activity list

|  |  |
| --- | --- |
| Date | 03 November 2022 |
| Team ID | PNT2022TMID04005 |
| Project Name | Project -Detecting Parkinson’s Disease using Machine Learning |

# Data Collection

1. Download a dataset.

# Image Pre-Processing

1. Importing the necessary Libaries
2. Loading Train data and test data
3. Quantifying images
4. Label encoding

# Model Building

1. Training the model
2. Testing the model
3. Model Evaluation
4. Save the model

# Application Building

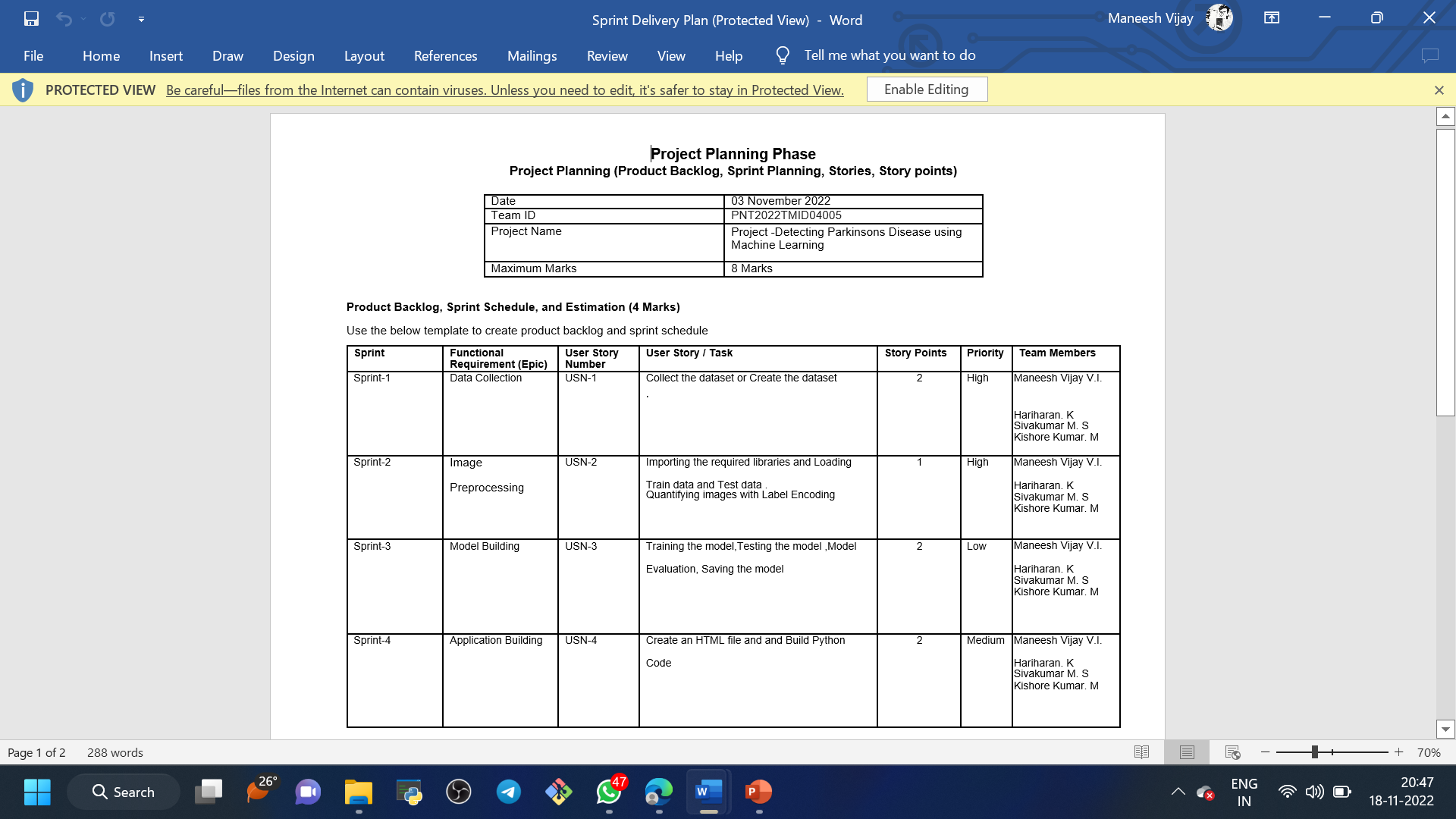
1. Building HTML pages
2. Build Python code
3. Run the app

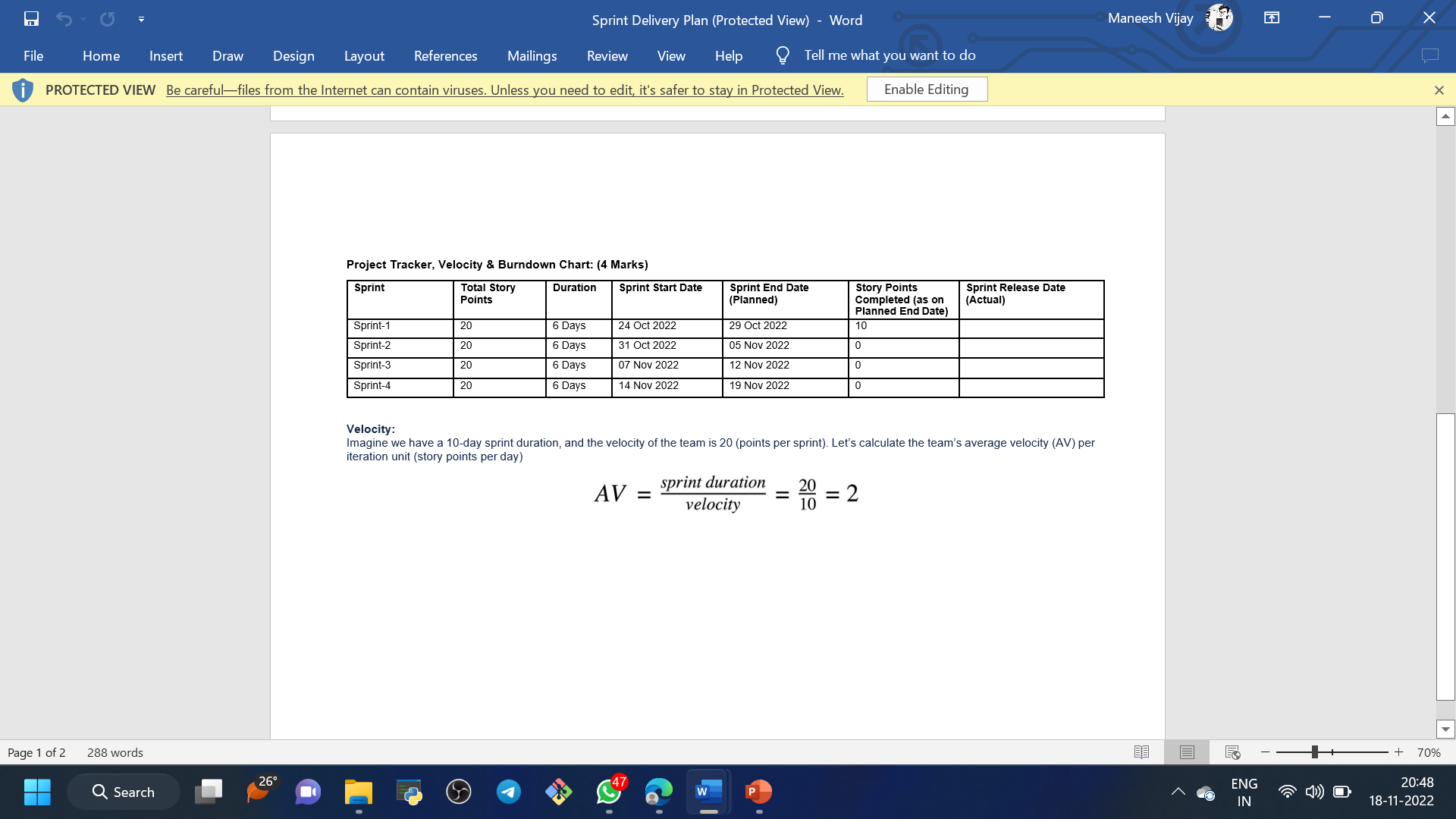
# Train the model on IBM

# Register for IBM cloud

1. Train model on IBM Watson

**SPRINT DELIVERY**





**6.CONCLUSION:**

Parkinson’s disease affects the CNS of the brain and has yet no treatment unless it’s detected early. Late detection leads to no treatment and loss of life. Thus its early detection is significant. For early detection of the disease, we utilized machine learning algorithms such as XGBoost and Random Forest. We checked our Parkinson disease data and find out XGBoost is the best Algorithm to predict the onset of the disease which will enable early treatment and save a life.