Machine learning Applications in Healthcare

*Submitted in partial fulfilment of the requirements*

*of the degree of*

BACHELOR OF ENGINEERING

*in*

INFORMATION TECHNOLGY

(A.Y. 2018-2019)

by

|  |
| --- |
| Varun Mishra (Roll No.: 47)  Darpan Mehra (Roll No.: 42) |

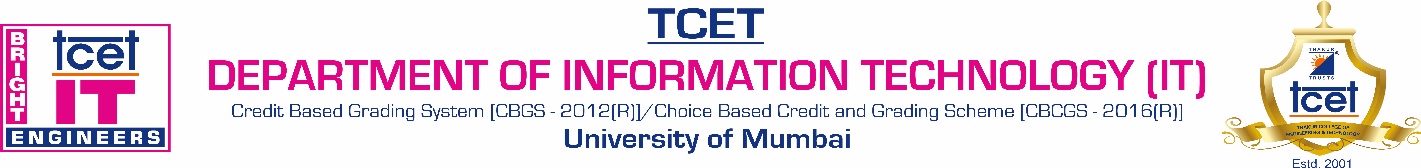
Under the Guidance of

Mr. Shridhar Kamble

Designation, I.T Department, TCET



**University of Mumbai**



**Certificate**

This is to certify that Mr./Ms.\_\_\_\_\_\_\_\_\_, Mr./Ms.\_\_\_\_\_\_\_, Mr./Ms.\_\_\_\_\_\_\_\_\_,

are bonafide students of Information Technology Department, Thakur College of Engineering and Technology, Mumbai. They have satisfactorily completed the requirements of PROJECT-I as prescribed by the University of Mumbai, while working on “Project Title”.

|  |  |
| --- | --- |
| Signature :------------------------------Name : Internal Guide’s Name  Designation | Signature :------------------------------  Name : Dr. Rajesh S. Bansode  HOD-IT |
| Signature: ---------------------------------------------  Name : Dr. B. K. Mishra  Principal,  Thakur College of Engineering and Technology. | |

|  |
| --- |
| **Internal Examiner: External Examiner:** |
| Signature :--------------------------------- Signature :---------------------------------  Name : Name : |
| Thakur College of Engineering and Technology, Kandivali(East) Mumbai.  Date:  Place: |

**Declaration**

I/we declare that this written submission represents my/our ideas in my/our own words and where others ideas or words have been included, I/we have adequately cited and referenced the original sources. I/we also declare that I/we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my/our submission. I/we understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

----------------------------------------- (Signature)

1.---------------------------------------------

2.---------------------------------------------

(Name of studentandRoll No.)

Date:

**ACKNOWLEDGEMENT**

We sincerely thank to our guide Mr./Ms.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

for his/her guidance and support for carrying out our project work.

1.---------------------------------------------

2.---------------------------------------------

(Name of studentandRoll No.)

**Note:** While writing the acknowledgement all students are supposed to follow the hierarchy in the order of principal, Mentor Dean, HOD, guide, colleague, parents as applicable.

**ABSTRACT**

Machine learning is a fairly new technology, with new developments coming up faster than most people can learn them. It is a cutting-edge technology in every sense of the word. In such a case, it is necessary that we handle it with care. It could very well throw up results at an accuracy rate that is infeasible for medical applications. The capabilities required for the project are fairly in the feasible range. Most of the planned feature hinge around processing datasets. Due to the above, a challenge also visible here is the problem of dealing with huge amounts data and processing it thoroughly to get proper results.

The central parameter of success is the accuracy at which the machine learning algorithm predicts the diseases for which it is trained. It is expected to be above 96% - 98% to make a real difference, as medical diagnosis has very rigorous standards. The secondary parameters are the response and loading time of each modules. A high value of module latency can make any user impatient. Besides these 2, the other parameter is the number of features that work without any bugs. This parameter should be ideally above 90% as well, or it can make users frustrated.

Although machine learning is a nascent specialization, we found an abundance of algorithms, not just in the proposed state, but rigorously scrutinized in research papers. Convolutional neural networks were suggested for their high accuracy. Chinese researchers had used Convolutional neural networks in a large-scale analysis of medical records of hospitals in China and reported high accuracy in prediction through symptoms [2]. Besides this, many researchers had tested SVM and ANNs, which were conventionally image classification algorithms. They also reported high accuracy [4].

Until now, our work has focused on the training of our algorithm, with the use of medical data available openly. The algorithm has not shown the required accuracy and will need to be trained further. The other module we have built is the GUI. We built a web-based GUI in flask, an open source python framework. These two modules are the endpoints of the project and building them first allows the ones in the middle to be built with convenience.

The comparison between existing results and actual results show a gap in accuracy that has to be closed. However, we were able to achieve more versatility, which was probably the reason that the algorithm was less accurate in the first place.

**C O N T E N T S**

* + - * List of Figures i
      * List of Tables ii
      * Abbreviations and symbols iii

**ChapterNo. Topic Pg. No.**

**Chapter1 Overview 01**

**1.1 Introduction**

1.2 Background

1.3 Importance of the Project

1.4 Perspective of stakeholders and customers

1.5 Objectives and Scope of the project

**1.6 Summary**

**Chapter2 Literature Survey & Proposed Work (Phase wise) 05**

**2.1 Introduction**

**2.2 Literature Survey Table**

2.3 Problem definition (Phase wise)

2.4 Feasibility Study

2.5 Methodology used

2.5.1 Scrum/XP/Agile

2.5.2 Customer interaction details

**2.6 Summary**

**Chapter3 Analysis and Planning**

**3.1 Introduction**

3.2 Product Backlog or Sprint backlog

3.3 Project planning (Resources, Tools used, etc.)

3.4 Scheduling (Time line chart or Gantt chart) according to sprint backlog

**3.5 Summary**

**Chapter4 Design and Implementation**

4.1 DFD

4.2 Flow Chart

4.3 UML

4.4 GUI screenshot

4.5 Database screenshot

**Chapter 5 Results & Discussion**

5.1Actual Results

a. Outputs (sprint wise)

b. Outcomes

c. Discussion of the results

5.2 Future Scope (further phases)

5.3 Testing

5.4 Deployment

**Chapter 6 Conclusion**

6.1 Conclusion

**Chapter 1 Overview**

1.1 Introduction

Healthcare, though a new venture for the Artificial intelligence\ Machine learning domain, remains one of the most crucial domains of public service, one which has been devoted a lot of funding and research. With healthcare, we are possibly looking at the most cardinal and relevant applications of the advanced abilities of Artificial intelligence as well as Machine learning.

With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care, and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks.

In such a case, it is necessary that we utilize mathematical modelling to iron out any imperfections and aberrations, so that we have the rigorous standards of accuracy that the medical diagnosis field requires.

Artificial intelligence presents challenges due the complexity involved in getting the balance between too much and not enough. We may design systems which can process an extremely wide variety of inputs, but we cannot actively ensure that the AI will respond to it in the way we want it to. It may present us an output which was intended to be for another input, or it may not be able to process it due to the load caused by the heavy processing on the interpreter. Thus, the challenge is not just of input handling or building features, but of efficiency as well.

1.2 Background

Many virtual assistants are available on the market, and they give high accuracy in handling input, particularly Siri and Google Assistant.

Siri is a virtual assistant part of Apple Inc.'s iOS, watchOS, macOS, HomePod, and tvOS operating systems. The assistant uses voice queries and a natural-language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Internet services. The software adapts to users' individual language usages, searches, and preferences, with continuing use.

Google Assistant is a virtual assistant powered by artificial intelligence and developed by Google that is primarily available on mobile and smart home devices. Unlike Google Now, the Google Assistant can engage in two-way conversations.

Amazon Alexa is a virtual assistant developed by Amazon, first used in the Amazon Echo and the Amazon Echo Dot smart speakers developed by Amazon Lab126. It is capable of voice interaction, music playback, making to-do lists, setting alarms, streaming podcasts, playing audiobooks, and providing weather, traffic, sports, and other real-time information, such as news. Alexa can also control several smart devices using itself as a home automation system.

Although not anywhere close in accuracy, the application understands most commands with enough accuracy to extract results through them. Many assistants exist in the healthcare domain too. However, not many of them offer a catalogue of services along with Machine learning applications.

Besides these assistants, there are systems without AI capabilities as well, such as WebMD, which provide symptom matching without the use of AI features.

While one does come across a lot of AI steeped healthcare software, what one doesn’t see often is these services consolidated as a single package. For users, especially patients on medication or bedrest, it is paramount that they receive care without the vexation of having to juggle between multiple software platforms, annoyed as they already are at having to juggle between medicines and health routines.

1.3 Importance of the project

The project aims at creating an intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other hospital staff work efficiently. This also aims at having a basic machine learning capability for disease prediction.

In an era where medical diagnosis and treatment has made massive leaps, it is unfair to deny anyone proper healthcare owing to a lack of means. This project aims to make healthcare less exclusive than it currently is and make it easier to find and avail services.

One often sees a lack of a proper software to help it out in such cases. Software can avoid the necessity of having to revisit doctors, or to spend hours self-researching complex medical terms and appendices, as many diagnosed patients tend to do. The use of a Bot here can automate such mundane tasks and still leave the patient satisfied and assured.

1.4 Perspective of stakeholders and customers

Stakeholder perspective is crucial to the success of this project. Part of the reason is that medical diagnosis is a highly sensitive field, and even the slightest of errors, which are evidently unavoidable in even the most sophisticated software, can lead to the patient’s condition worsening. It is also vital to involve the patients, as the patients being comfortable with the equipment, both hardware and software, is necessitous to the smooth working of the technology.

Stakeholders here include

* Doctors and nurses
* Health insurance companies
* Hospitals
* Health department (State as well as National/Federal)
* Biotechnology manufacturers

Doctors and nurses are highly conservative when it comes to adoption of technology, especially software. Due to the sensitivity of the data as well as the need for accuracy, most doctors do not adopt or advise any software, or rather any new technology, until it has been rigorously tested and reviewed by various independent evaluators.

Health insurance companies are much similar to doctors and nurses in this aspect, as there is an extremely large amount of money involved. However, customers looking for health insurance plans were eager for such kinds of technology, evident by the high amount of Google Searches for health insurance plan choosers.

Hospitals were more amenable as well, provided they were supported by the doctors. The use of bleeding edge technology in medical diagnosis is something that many hospitals in California, Massachusetts and New York were willing to try out. Many had, in fact, already done large scale tests on patients.

Health departments have a lot of restrictions on such technology, unless it is being used in a non-invasive way or is complementary to orthodox methods. There are numerous regulations concerning the use of cutting edge technology in the medicinal field.

Biotechnology manufacturing concerns the manufacturing of diagnosis equipment, measuring equipment, support equipment, surgery tools etc. Manufacturers are usually enthusiastic about such technologies. It is also imperative to obtain their support as it makes synchronization between the devices easier.

Potential users, such as those who frequent the forums related to healthcare, wearable technology etc., seemed open to the idea, as anticipated. They seemed to like the idea of a machine learning being used to predict diseases, as well as the other minor features.

1.5 Objectives and Scope of the project

The project aims at creating an intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other hospital staff work efficiently. This also aims at having a basic machine learning capability for disease prediction.

The main objective is to offer a way for doctors and patients alike to easily perform mundane healthcare activities as well as advanced prediction services.

We plan to offer various services parallelly. These services include but are not limited to

1. Hospital finder

2. Record keeper

3. Symptom matching through machine learning

4. Insurance selector

5. Wearable technology manager

6. Information catalogue

7. Doctor finder

Besides the features above, we also aim to make the software robust, accurate and fast.

1.6 Summary

Healthcare, though a new venture for the Artificial intelligence\ Machine learning domain, remains one of the most crucial domains of public service. With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care, and community services.

It is necessary that we utilise mathematical modelling to iron out any imperfections and aberrations, so that we have the rigorous standards of accuracy that the medical diagnosis field requires.

Many virtual assistants are available on the market, and they give high accuracy in handling input, particularly Siri and Google Assistant. Although not anywhere close in accuracy, the application understands most commands with enough accuracy to extract results through them.

The project aims at creating an intelligent but lightweight healthcare assistant. In an era where medical diagnosis and treatment has made massive leaps, it is unfair to deny anyone proper healthcare owing to a lack of means. This project aims to make healthcare less exclusive than it currently is and make it easier to find and avail services. This application will be able to help healthcare and other hospital staff work efficiently.

Stakeholder perspective is crucial to the success of this project. It is also vital to involve the patients, as the patients being comfortable with the equipment, both hardware and software, is necessitous to the smooth working of the technology. Potential users seemed open to the idea, as anticipated.

Stakeholders here include - Doctors and nurses, Health insurance companies, Hospitals, Health department (State as well as National/Federal), Biotechnology manufacturers.

This application will be able to help healthcare and other hospital staff work efficiently. This also aims at having a basic machine learning capability for disease prediction.

The main objective is to offer a way for doctors and patients alike to easily perform mundane healthcare activities as well as advanced prediction services.

**Chapter 2 Literature Survey & Proposed Work (Phase wise)**

2.1 Introduction

Analysis of the current market scenario and technological capabilities is essential to the building of a project. Here we analysed various prevalent research papers pertaining to our project as a part of the literature survey. The project aims at creating an intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other hospital staff work efficiently. This also aims at having a basic machine learning capability for disease prediction.

Our literature survey was focused on the capabilities of the choices we made for the core part of the project, namely artificial neural networks and the use of wearable technology for data collection.

The initial phases focused on the planning and design of the project. We analysed extant products, customer interests and so on. For analysis, we looked at the current iterations of popular and cutting-edge algorithms used in machine learning, such as artificial neural networks and convolutional neural networks. The analysis was openly available in numerous research papers. The design involved comparing the project with some similar projects.

2.2 Literature Survey Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ref.  No. | Year | Author | Key findings (Results) | Research gaps |
| 2 | April 2017 | Chen, hao, hwang, wang, wang | 1) Big data has grown in healthcare, leading to potential for accuracy in medical analysis.  2) Wearable tech provides new ways to obtain data from patients and monitor their health.  3) A convolutional neural network was used to extract text characteristics.  4) 20 million records from 31 thousand patients were used in the analysis. Records were of three types   * Structured data * Text data * Structured and text   5) The text data is represented in vector format i.e. Word embedding.  6) The multi-model disease risk prediction algorithm gives an accuracy of 94.8%. | 1) Application method for real life healthcare tasks.  2) Methods to improve data structuring for better accuracy. |
| 3 | January 2017 | Chen, Ma, Li, Wu, Zhang, Youn | 1) Washable smart clothing consists of sensors, electrodes, and wires.  2) Advanced methods of data retrieval are combined with advanced methods of data analytics and cognitive computing.  3) Body temperature sensor is put in the underarm seam; a set of ECG sensors are mounted on the chest, shoulders, and ribs; the myocardial sensor is embedded in the left part of the chest; and the SpO2 sensor is deployed on the triceps.  4) The data is dependent on the emotional state of the subject.  5) 2 layers of storage are used, the first on local devices and the second on the high-level clouds for advanced processing. | 1) A proper method to collect data without making the user uncomfortable.  2) Usage of algorithms to handle data processing and compression. |
| 4 | March 2015 | S. Vijayarani, S.Dhayanand | 1) The healthcare industry collects huge amounts of healthcare data  2) This data was used to train a system running SVM and ANN algorithms. | 1) Methods to improve classification accuracy |
| 5 | November 2016 | Abadi, Barham, Chen, Dean, Devin, Warden, Ghemawat, Irving, Isard, Kudlur, Chen, Monga, Moore, Murray, Davis, Vasudevan | 1) Tensorflow is a large-scale, heterogeneous machine learning system, which uses dataflow graphs to represent computation.  2) Tensorflow is a descendant to DistBelief, which Google has used since 2011  3) Many neural networks are trained using stochastic gradient descent  4) Tensorflow works across a distributed GPU architecture to maximize efficiency. | 1) Methods to improve classification accuracy |
| 6 | June 2016 | Yong-Young Kim, Mi-Hye Kim | 1) Software developers & medical experts believe that health monitoring and the prevention of diseases should be explored.  2) Absence of certifications for healthcare software is still a concern | 1) Societal acceptance of automized healthcare  2) Specification of anticipated problems |
| 6 | October 2014 | Jürgen  Schmidhuber | 1) A standard neural network (NN) consists of many simple, connected processors called neurons  2) Re-cyclic neural networks are the deepest of all neural networks | 1) Constraints in application |
| 7 | November  2014 | Kouroua,  Exarchos,  Exarchos,  Karamouzis, Fotiadis | 1) Machine learning techniques are being utilized to model the progression and treatment of cancerous conditions.  2) It uses methods like Artificial neural networks, Support vector machine, Decision trees and Bayesian networks. For evaluating the performance, it uses Holdout method, random sampling, Cross validation and Bootstrapping. | 1) Application method for real life cancer detection.  2) Validation accuracy of algorithm. |

2.3 Problem definition (Phase wise)

We have divided the project into 2 phases, which are further divided into 5 and 2 sub-phases each. The problem definitions for each phase and sub-phase are as follows.

Phase I – Phase I will focus on forming a base for the project. With thorough planning, analysis and design, we will ensure that the actual implementation is smoother.

1. Planning: Applying agile methodology for planning our healthcare software project, and to achieve better results in the time frame given to us, along with better flexibility.
2. Analysis: Prepare an analysis of our own project. Prepare a detailed analysis on present healthcare assistants and overcoming its limitation and performance benchmarks.
3. Design: Integration of data and designing of event system app.
4. Coding: Writing the whole system app code and taking help of open source.
5. Implementation: Giving user to test alpha testing and gathering centric analysis of performance, feedback, and try to improve the quality of the result/output

Phase II – Phase II will focus on fine tuning the project and ensuring that the intended features will work as we wanted them to.

1. Testing: Doing the various test on system app like as Unit testing, Integration testing, Regression testing, system testing, etc., of test case data to check if the Integrated system functions are as desired by the client.
2. Deployment: Conduct beta testing for identifying any further errors, bugs and improvements that can be performed. After testing and approval, deploy the proposed system.

2.4 Feasibility study

Economic feasibility

The project relies on open source software, such as python, R, NumPy and SciPy. The functionality provided by these software’s is enough to construct a product capable of usage for healthcare applications.

For the machine’s learning and training, we will use freely available datasets. These sets run into thousands of lines, and thus can provide enough learning capability to the software. Due to these 2 being freely available, the project is very low cost and thus feasible from an economic standpoint.

The expected total cost is thus 0 rupees. As such, we do not need any outside funding. The project will be expected to make minimal sales, due to the presence of a variety of other similar tool and the slow adaption of software’s by the medical community.

Technical

The capabilities required for the project are fairly in the feasible range. Most of the planned feature hinge around processing datasets. A GUI is required, which will be done with the pyGUI framework. The GUI will be kept simple and minimalist. For the hospital searching and insurance selection, we will use readily available data. We also have a feature for users to set personal data, preferences.

For the machine learning, we will use Neural networks to train our datasets, both of which are available on open source repositories.

Operational

On an operational level, the software will need to run intensive processing tasks while it’s in its learning phase. Since the learning phase is a very short process, it is doable without the need for extra hardware. The software will be deployed as a Desktop application and will need users to download it. It will also need internet to operate.

Socio-cultural

This application will have many positive applications on society. It will enable users to manage their healthcare needs in a cheap and easy way. Users with chronic illnesses will find it especially helpful. Those with financial problems will also find it helpful.

Some users may not be willing to accept automated healthcare, but most would be willing. Since it reduces the need for health monitoring, users living in areas with costly healthcare (e.g. most of the USA) will find it helpful. Also, the sharing of instantly available, accurate information can make the difference between life and death.

Legal

Since the project deals with a sensitive topic with potential for major harm, it will carry a disclaimer for improper usage. The software is not meant to be a substitute for doctors or hospital provided healthcare. Users should ideally only use it to supplement professional healthcare. However, the potential for misuse, whether intentional or not, still exists.

Much has been said about the ethical advantages of free software in general, and it is particularly true in a profession in which the sharing of instantly available, accurate information can make the difference between life and death. As medical software begins to offer decision support, risk management, performance rating, and analytic features, physicians should learn to be more careful with its usage.

Ethical

While the software is meant to help users with their health, there still is potential for improper usage leading to damage. The software should not be used as a substitute for doctors or professionally provided healthcare. Users must only use it to supplement the same. Much has been said about the ethical advantages of “free” software in general, and it is particularly true in a profession in which the sharing of instantly available, accurate information can make the difference between life and death.

As medical software begins to offer decision support, risk management, performance rating, and analytic features, physicians should not accept black boxes and secret formulas that constrain sharing and intimately affect patient care and reimbursement.

2.5 Methodology used

The methodology that we chose for our project is Agile. While waterfall model is a tried and tested framework, it is not suitable for our project, due to the changeable nature and small team size. Agile allows us to be flexible, manage changing requirements, manage the ever-increasing scope as well as get consumer perspective.

Agile has focus on customer satisfaction, which is a crucial aspect of any healthcare application. Patients, insurance providers and doctors need to be clued in to the process for the indispensable inputs they can provide.

2.5.1. Agile

Agile software development is a methodology for software projects under whose framework project requirements and solutions evolve through the collaborative effort of self-organizing and cross-functional teams and their customer and intended users. It advocates for flexible modelling and planning, flexible implementation, rapid delivery of code, and constant improvement, and it advocates speedy and flexible reactions to change.

12 Agile Principles

1. Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3. Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. Business people and developers must work together daily throughout the project.
5. Build projects around motivated individuals. Give them the environment and support they need and trust them to get the job done.
6. The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
7. Working software is the primary measure of progress.
8. Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
9. Continuous attention to technical excellence and good design enhances agility.
10. Simplicity - The art of maximizing the amount of work not done--is essential.
11. The best architectures, requirements, and designs emerge from self-organizing teams.
12. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

2.5.2 Customer interaction details

We used forums as the main source of our customer input. AI forums as well as healthcare forums are buzzing with activity. Users showed interest in Machine Learning based prediction of diseases, as well as a software that can do it for a reasonable cost.

Besides this, we researched interests of customers in wearable technologies. Technologies like Fitbit have already made a splash in the healthcare market. Users of Fitbit displayed interest in software products that could help them better utilize the data that it was gathering.

2.6 Summary

Here we analyzed various prevalent research papers pertaining to our project as a part of the literature survey. The project aims at creating an intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other hospital staff work efficiently.

Our literature survey was focused on the capabilities of the choices we made for the core part of the project, namely artificial neural networks and the use of wearable technology for data collection.

We have divided the project into 2 phases, which are further divided into 5 and 2 sub-phases each. The problem definitions for each phase and sub-phase are as follows.

Phase I – Phase I will focus on forming a base for the project. With thorough planning, analysis and design, we will ensure that the actual implementation is smoother.

Phase II – Phase II will focus on fine tuning the project and ensuring that the intended features will work as we wanted them to.

Due to these 2 being freely available, the project is very low cost and thus feasible from an economic standpoint.

The expected total cost is thus 0 rupees. As such, we do not need any outside funding.

The capabilities required for the project are fairly in the feasible range. Most of the planned feature hinge around processing datasets. A GUI is required, which will be done with the pyGUI framework. The GUI will be kept simple and minimalist. For the hospital searching and insurance selection, we will use readily available data. We also have a feature for users to set personal data, preferences. On an operational level, the software will need to run intensive processing tasks while it’s in its learning phase.

This application will have many positive applications on society. It will enable users to manage their healthcare needs in a cheap and easy way. Since the project deals with a sensitive topic with potential for major harm, it will carry a disclaimer for improper usage. The software is not meant to be a substitute for doctors or hospital provided healthcare. Users should ideally only use it to supplement professional healthcare. However, the potential for misuse, whether intentional or not, still exists.

The methodology that we chose for our project is Agile. Agile allows us to be flexible, manage changing requirements, manage the ever-increasing scope as well as get consumer perspective. Agile has focus on customer satisfaction, which is a crucial aspect of any healthcare application.

Agile software development is a methodology for software projects under whose framework project requirements and solutions evolve through the collaborative effort of self-organizing and cross-functional teams and their customer and intended users.

Users showed interest in Machine Learning based prediction of diseases, as well as a software that can do it for a reasonable cost.

Besides this, we researched interests of customers in wearable technologies. Technologies like Fitbit have already made a splash in the healthcare market.

**Chapter 3 Analysis and planning**

3.1 Introduction

The analysis phase is the most crucial phase of any project. The quality of the analysis can make or break a project. Artificial intelligence presents challenges due the complexity involved in getting the balance between too much and not enough. Planning such a task is matter of finding what we really need and we don’t.

We may design systems which can process an extremely wide variety of inputs, but we cannot actively ensure that the AI will respond to it in the way we want it to. It may present us an output which was intended to be for another input, or it may not be able to process it due to the load caused by the heavy processing on the interpreter. We have to plan for such a scenario too. Thus, the challenge is not just of input handling or building features, but of efficiency as well.

We did this phase slowly so we be could be thorough with all our requirements and plans. A slow approach also allowed us to anticipate risks. Planning phase is probably the best time to plan for risks and avoid them altogether. Being careful in the planning phase allows us to put less effort into the risk mitigation, risk monitoring and risk management plan.

3.2 Product Backlog or Sprint backlog

The backlog of the project is given as follows:

To do

* Obtain insurance data
* Build an insurance selector
* Make GUI for fringe features
* Add a hospital finding feature
* Add a symptom matching feature
* Add an archiving feature
* Make a checklist database for users
* Obtain a fitness plan database
* Obtain global data for patients to compare with

Ongoing

* Information catalogue building
* Dataset training
* GUI building
* Procurement of more medical data for higher accuracy

Done

* Analysis
* Design
* Modelling
* Literature survey
* Budgeting and scheduling
* Feasibility analysis

3.3 Project planning (Resources, Tools used, etc.)

IT projects require resources in terms of money, time, human resources, infrastructure and technology, both hardware and software. Resources are not just a mean, but also an approximation of constraints.

Project planning is essential to managing the scope, schedule and budget of the project. For this, we used tools such as MS Excel, MS PowerPoint, online MS project as well as various modelling tools, such as draw.io.

Stakeholder perspective is crucial to the success of this project. Part of the reason is that medical diagnosis is a highly sensitive field, and even the slightest of errors, which are evidently unavoidable in even the most sophisticated software, can lead to the patient’s condition worsening.

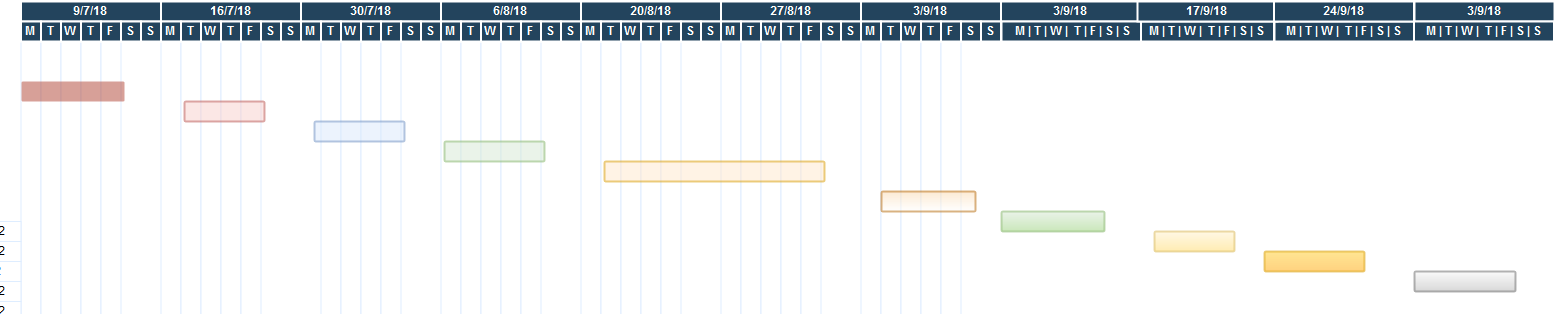
Thus, we made it a necessity to search for user consensus before we planned for features to be built in our project. This was done by researching search interest on search engines, visiting forums pertaining to machine learning, artificial intelligence, healthcare as well as wearable technology.

We divided our project into various phases and sub-phases, and allocated date ranges from a week to 3 weeks to every sub-phase. This was done using the timeline chart feature of online MS project. We then used a Gannt chart to model schedule dependencies and fine tune the scheduling. The Gantt chart was created using the Gantt chart feature available in MS Excel. Using these two, we further planned our sub-phases.

The modelling included charts and diagrams such as UML diagram, GUI modelling, dependencies and other diagrams. These were done using free tools like draw.io. Drawing these models helped to incisively assess our requirements and features.

3.4 Scheduling (Time line chart or Gantt chart) according to sprint backlog

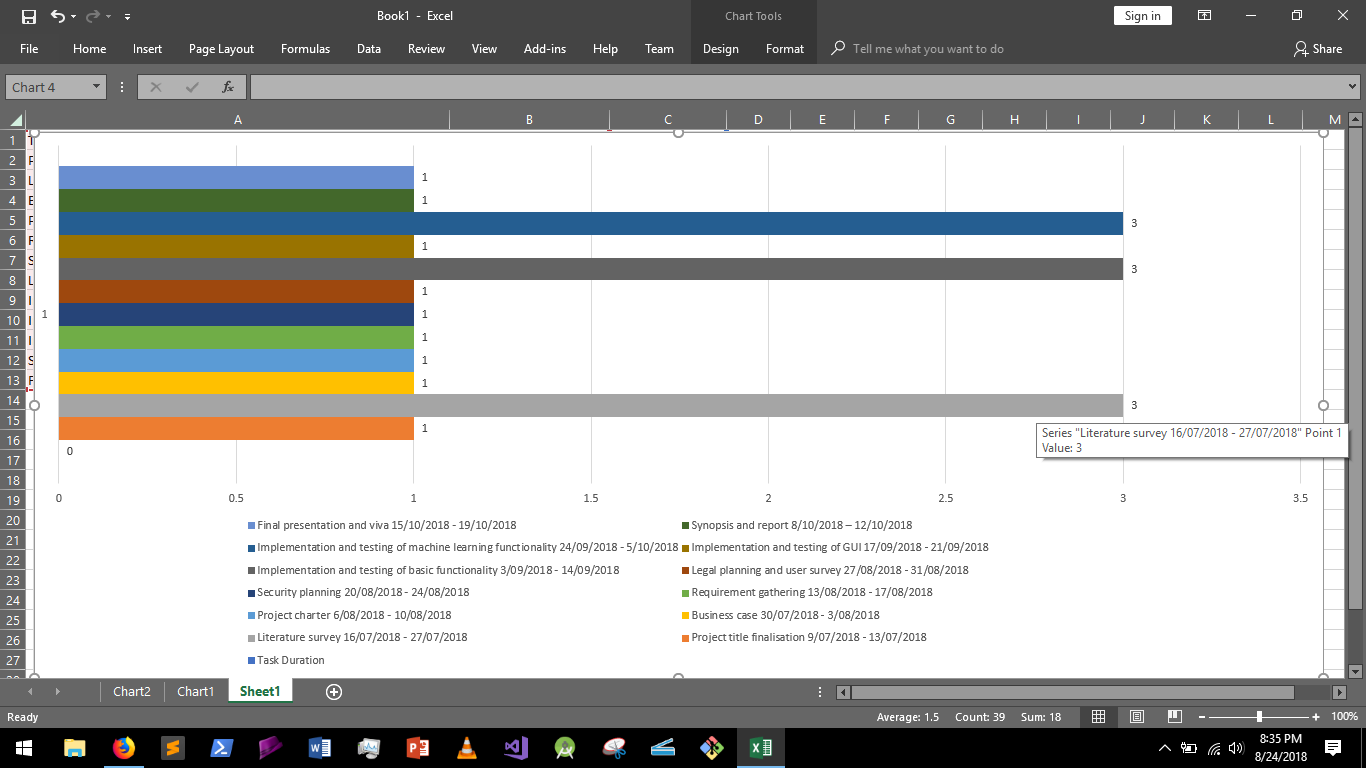
|  |  |  |
| --- | --- | --- |
| **Task** | **Duration** | **No. of days** |
| Project title finalization | 9/07/2018 - 13/07/2018 | 1 |
| Literature survey | 16/07/2018 - 27/07/2018 | 3 |
| Business case | 30/07/2018 - 3/08/2018 | 1 |
| Project charter | 6/08/2018 - 10/08/2018 | 1 |
| Requirement gathering | 13/08/2018 - 17/08/2018 | 1 |
| Security planning | 20/08/2018 - 24/08/2018 | 1 |
| Legal planning and user survey | 27/08/2018 - 31/08/2018 | 1 |
| Implementation and testing of basic functionality | 3/09/2018 - 14/09/2018 | 3 |
| Implementation and testing of GUI | 17/09/2018 - 21/09/2018 | 1 |
| Implementation and testing of machine learning functionality | 24/09/2018 - 5/10/2018 | 3 |



Gantt Chart

|  |  |  |
| --- | --- | --- |
| **Task** | **Duration** | **No. of days** |
| Project title finalization | 9/07/2018 - 13/07/2018 | 1 |
| Literature survey | 16/07/2018 - 27/07/2018 | 3 |
| Business case | 30/07/2018 - 3/08/2018 | 1 |
| Project charter | 6/08/2018 - 10/08/2018 | 1 |
| Requirement gathering | 13/08/2018 - 17/08/2018 | 1 |
| Security planning | 20/08/2018 - 24/08/2018 | 1 |
| Legal planning and user survey | 27/08/2018 - 31/08/2018 | 1 |
| Implementation and testing of basic functionality | 3/09/2018 - 14/09/2018 | 3 |
| Implementation and testing of GUI | 17/09/2018 - 21/09/2018 | 1 |
| Implementation and testing of machine learning functionality | 24/09/2018 - 5/10/2018 | 3 |
| Synopsis and report | 8/10/2018 – 12/10/2018 | 1 |
| Final presentation and viva | 15/10/2018 - 19/10/2018 | 1 |

Bar representation



3.5 Summary

Here we summarize the analysis and planning phase of our project. The phase included scheduling, budgeting and creation of subtasks. The subtasks help us properly define our needs and features.

The analysis phase is the most crucial phase of any project. The quality of the analysis can make or break a project. Artificial intelligence presents challenges due the complexity involved in getting the balance between too much and not enough. Planning such a task is matter of finding what we really need and we don’t.

We did this phase slowly so we be could be thorough with all our requirements and plans. A slow approach also allowed us to anticipate risks.

The product backlog was divided into three phases - to do, ongoing and done.

The to do phase included - obtain insurance data, build an insurance selector, make GUI for fringe features, add a hospital finding feature, add a symptom matching feature, add an archiving feature, make a checklist database for users, obtain a fitness plan database, obtain global data for patients to compare with.

The ongoing tasks were - information catalogue building, dataset training, GUI building, procurement of more medical data for higher accuracy.

The tasks which were already done were - analysis, design, modelling, literature survey, budgeting and scheduling, feasibility analysis.

Project planning is essential to managing the scope, schedule and budget of the project. For this, we used tools such as MS Excel, MS PowerPoint, online MS project as well as various modelling tools, such as draw.io.

Stakeholder perspective is crucial to the success of this project. Part of the reason is that medical diagnosis is a highly sensitive field, and even the slightest of errors, which are evidently unavoidable in even the most sophisticated software, can lead to the patient’s condition worsening. Thus, we made it a necessity to search for user consensus before we planned for features to be built in our project.

The scheduling included a plan for the first half of the year. The plan includes - project title finalization, literature survey, business case, project charter, requirement gathering, security planning, legal planning and user survey, implementation and testing of basic functionality, implementation and testing of GUI, implementation and testing of machine learning functionality, synopsis and report and final presentation.

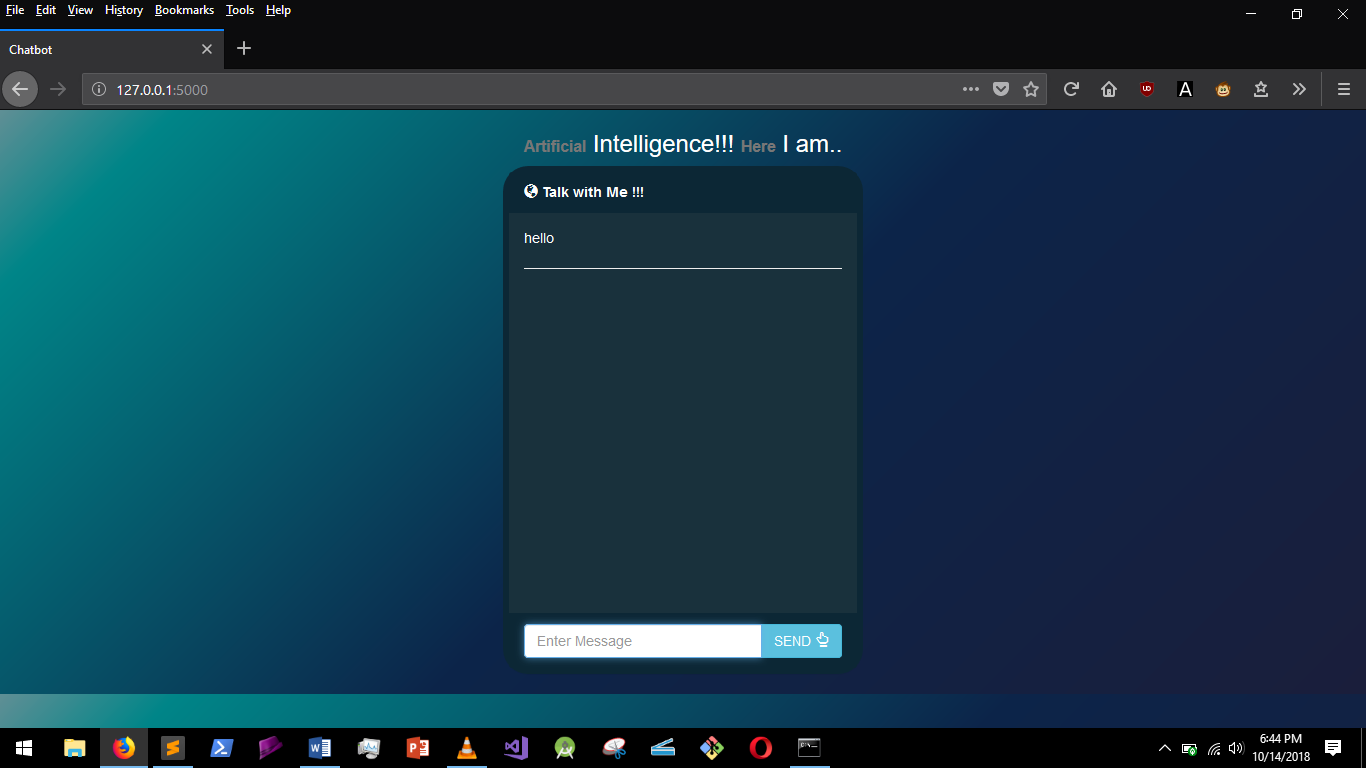
**Chapter 4 Design and Implementation**

4.1 DFD

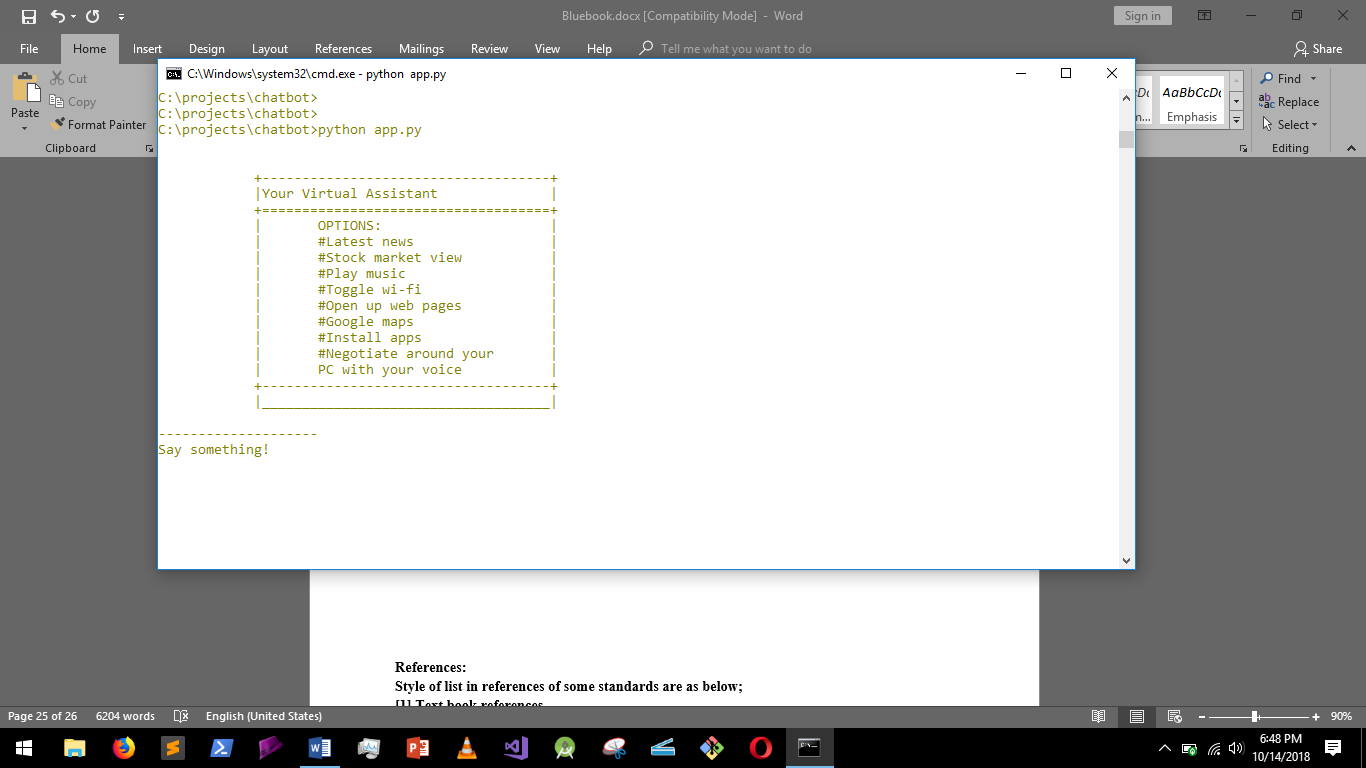
4.2 Flowchart

4.4 GUI

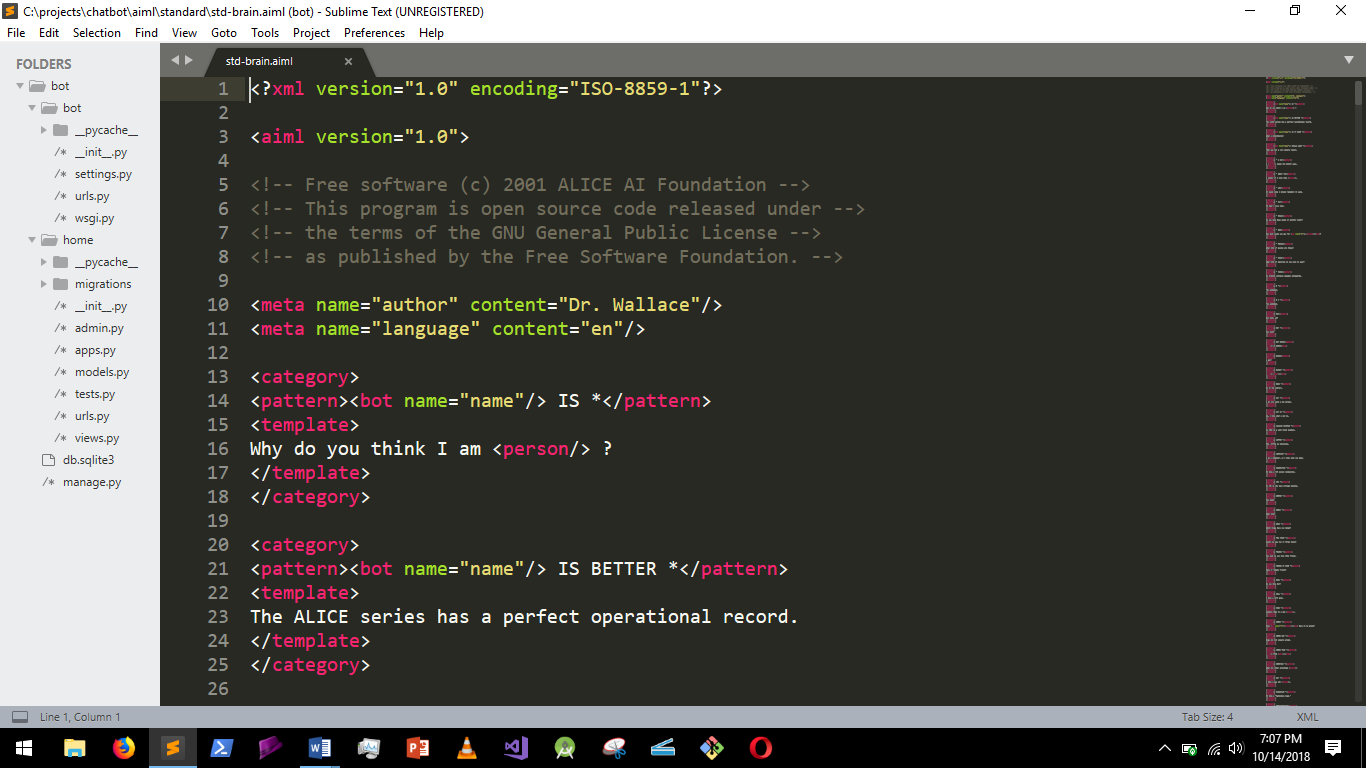
Web GUI

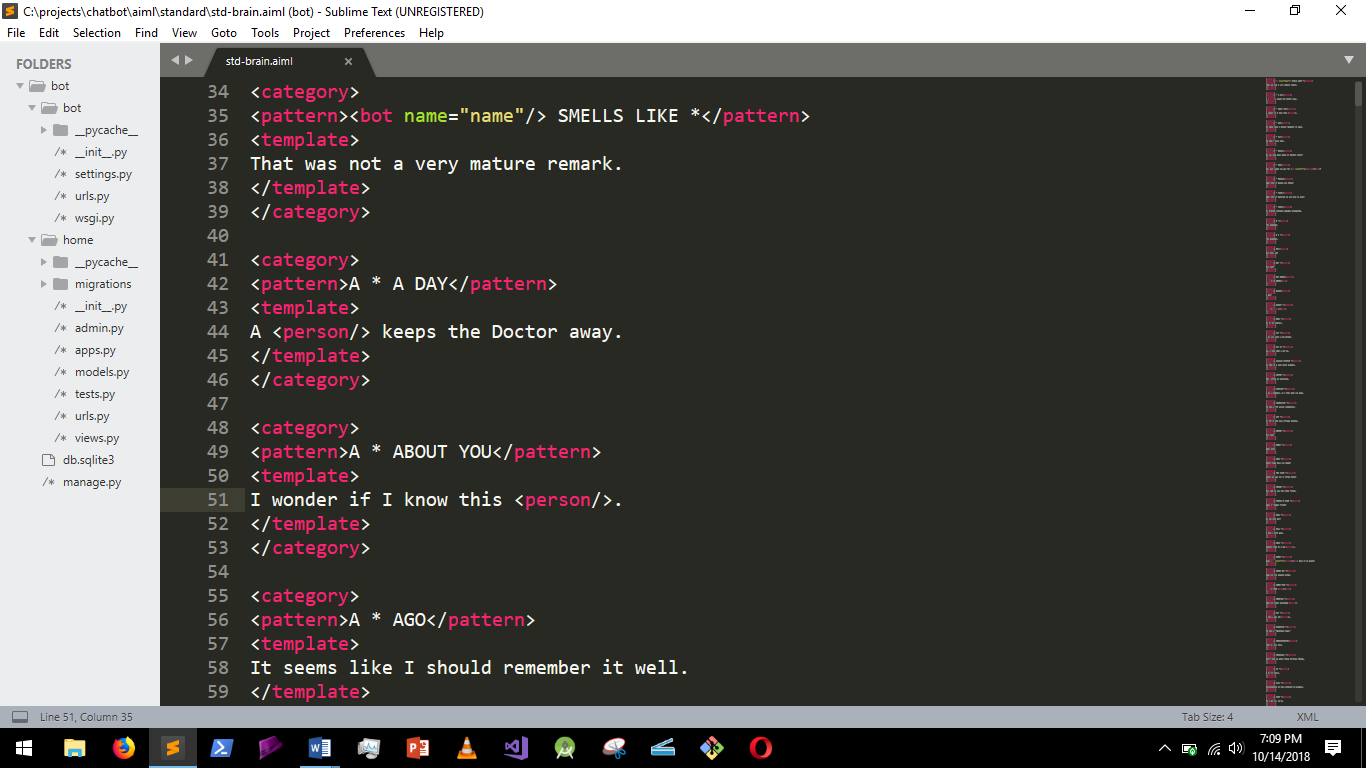


Command line

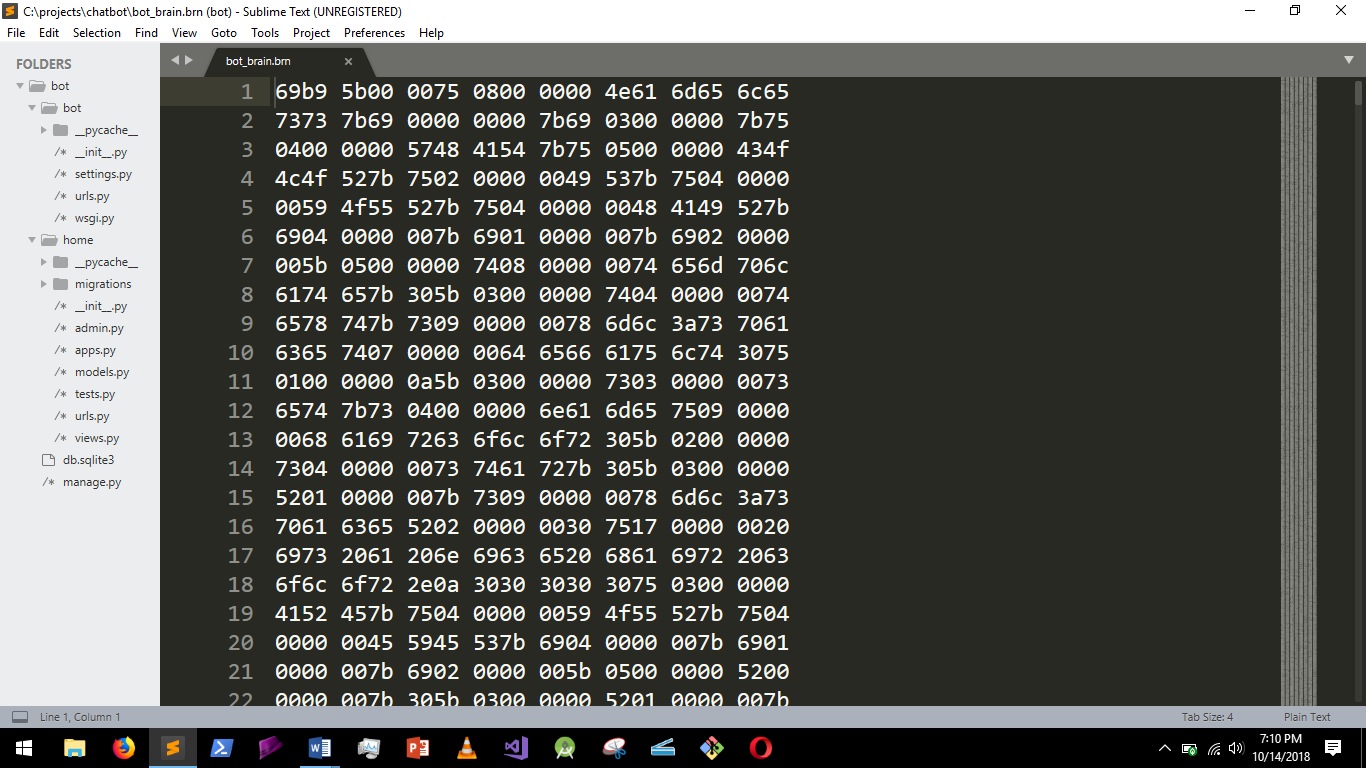


4.5 Database screenshots





Binary storage of data



**References:**

[1] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis, “Machine learning applications in cancer prognosis and prediction”, Computation and structural journal, Elsevier, 2014.

[2] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang, “Disease Prediction by Machine Learning Over Big Data from Healthcare Communities”, special section on healthcare big data, IEEE access, 2017.

[3] Min Chen, Yujun Ma, Yong Li, Di Wu, Yin Zhang, and Chan-Hyun Youn, “Wearable 2.0: Enabling Human-Cloud Integration in Next Generation Healthcare Systems IEEE Communications Magazine, 2017.

[4] Dr. S. Vijayarani, Mr. S. Dhayanand, “Kidney Disease Prediction Using SVM And ANN Algorithms”, International Journal of Computing and Business Research (IJCBR), 2015.

[5] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, Google Brain, “Tensorflow: A System for Large-Scale Machine Learning”, 12th USENIX Symposium on Operating Systems Design

and Implementation (OSDI ’16), 2016.

[6] Jürgen Schmidhuber, “Deep learning in neural networks: An overview”, Neural Networks Elsevier, 2014

[7] Yong-Young Kim and Mi-Hye Kim, “What are Software Developers’ and Medical Experts’ Priorities for Adopting a Healthcare Software Platform?”, Indian Journal of Science and Technology, 2016.

**[8]** [**https://en.wikipedia.org/wiki/Agile\_software\_development**](https://en.wikipedia.org/wiki/Agile_software_development)

**[9]** [**https://en.wikipedia.org/wiki/Amazon\_Alexa**](https://en.wikipedia.org/wiki/Amazon_Alexa)

**[10]** [**https://en.wikipedia.org/wiki/Siri**](https://en.wikipedia.org/wiki/Siri)

**[11]** [**https://en.wikipedia.org/wiki/Google\_Assistant**](https://en.wikipedia.org/wiki/Google_Assistant)

**[12]** [**http://agilemanifesto.org/principles.html**](http://agilemanifesto.org/principles.html)

**APPENDIX:**

**[A] Literature survey papers**

**[B] Plagiarism check report:**

1page plagiarism self- evaluation report. (Use Quetext or Plagscan for generating report)

**[C] Graduate Attributes and its mapping with the project**