INTRODUCTION

Objective: Predicting Disease from Symptoms

Building a web application based on ML model for predicting the type of disease based on the symptoms provided by the scans, x-rays and other medical data.

The app will provide the users with portability, instant access to necessary information and easier and quick communication with the doctors.

Data are gathered from various sources and repositories including Kaggle, UCI archive, etc, and supervised machine learning models are deployed to predict the disease outcome.

According to McKinsey Report
Use case: Predictive analytics in Healthcare sector

Doctor to Patient ratio

1:1445 INDIA

1:1000 WHO's prescribed norm

Primary Research

Age Group

57% Under- 35

43% Above- 35

Gender

65% Male

35% Female

Preference

72% Public Hospitals

28% Private Hospitals

Avg. Waiting Time

Public Hospitals 25 mins

Private Hospitals 15 mins

Avg waiting time for men is six times lower than that of women.

METHODOLOGY

Problem Structuring

- Building a web application based on machine learning model for predicting the type of disease based on the symptoms provided by the scans, synopsis, medical history and other data of the patients.
- Based on the nature of the disease the application will give recommendation for preventive measures and the nearest hospitals.

Data Collection & organization

- Collect the data from various sources including datasets from Kaggle, analyticsdimag, uci archive etc.
- Data are collected and organized into csv file and classified into broad categories – common diseases, heart diseases, chest X-ray, breast cancer diabetes, brain tumor, etc.
- Datasets are preprocessed for any imbalances and skewness.

ML model selection & implement

- Train the model using ensemble model with grid searching for hyperparameter tunning.
- Use supervised learning algorithm like naïve bayes, random forest, logistic regression and decision tree for classification of diseases.
- Use convolution neural network for training of image data like x-ray , mri scans.

Results & Recommendations

- Based on the analysis of scans, symptoms and medical history of the patient, the disease will be predicted by the web application.
- Recommendations about a nearby specialized doctor / hospital can also be given.
- The doctor profile can also be integrated to the app to facilitate a quick communication with the patients.

MARKET ANALYSIS

SWOT Analysis of Healthcare Industry in India

Strength
1) Elaborate network of Healthcare facilities
2) Location advantage of all healthcare buildings in the government sector.

Weakness
1) A skewed distribution of healthcare infrastructure.
2) Poor regulation and marketing of devices and equipment.
3) Long average waiting time in hospitals and poor doctor : patient

Opportunities
1) Commitment of govt to improve the present situation.
2) Improve the doctor : patient ratio.
3) Increasing domestic market for production of devices and equipment.

Threat
1) Govt is not able to maintain the present infrastructure.
2) Lack of funding to the healthcare sectors.
3) Political influence in the sector.

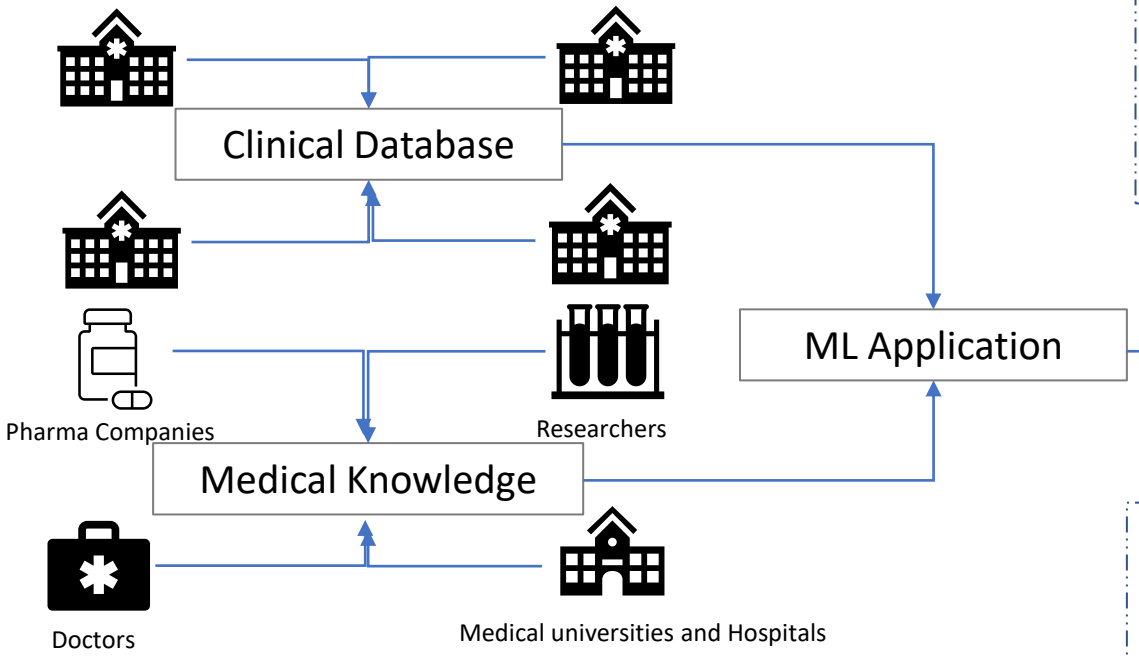
Customer (End user requirements):
Long average waiting time and non availability of centralized data are the major pain point for customers.

Target Audience/ Customer:
Doctors, patients, local pharma retailers, medical practitioners

Competitors in the market:
Roam Analytics, Sopris Health, Artelus, Sense.ly

Unique Selling point:
Helps in detection of possible disease in early stages, allowing for more effective treatment. Provides options for finding nearest hospitals , booking appointment.

PROPOSED SOLUTION



Input from User:
1) Age
2) Gender
3) Symptoms
4) Duration of symptoms/ past history
5) Location

Input from connected devices like gps, fitness devices, etc.

Output to User:
1) Type of disease
2) Preventive measures
3) Nearest Hospitals
4) Books an appointment with doctor

Technology driven value for users

1) **Effective Care Delivery:** Inputs from Connected device and fitness application are collected to get a holistic view of the disease of the patient.
2) **Viability and waste reduction:** Elimination of common low value procedures like high waiting time in hospitals. ML enabled waste, money and waiting time reductions.
3) **Enhanced Clinical productivity:** ML enabled clinical operations. Improving disease diagnosis and identifying gaps in patient care.
4) **Centralized database system:** With the help of consent manager mechanism, patients and hospitals can share, store and organized the data in a central repository.
5) **Recommendation:** The web app will give recommendation based on the inputs from the user. It will also access other parameters collected from other sources. Provide recommendation for nearest hospitals and an option to book an appointment with doctor.

Technology Landscape Assessment

We are working to make a Web App which collects different symptoms from people, make their medical profile of a person and try to predict if they are facing any possible disease. We have grouped diseases into 2 groups - Common diseases and Severe diseases (like Heart Disease, Cancer, Tumours, etc) using X-Rays and MRI scans.

So for model purposes, we are training the models using ensemble method with Grid search for hyperparameter running (which uses different algorithms and gives the one with best output). The algorithms being considered are Naive Bayes, Logistic Regression, Random forests and decision trees (Using ScikitLearn for data input and model training).

And for Image data for x-ray and mri, the input should be an image file and we'll be using convolution neural networks for model training (Using TensorFlow)

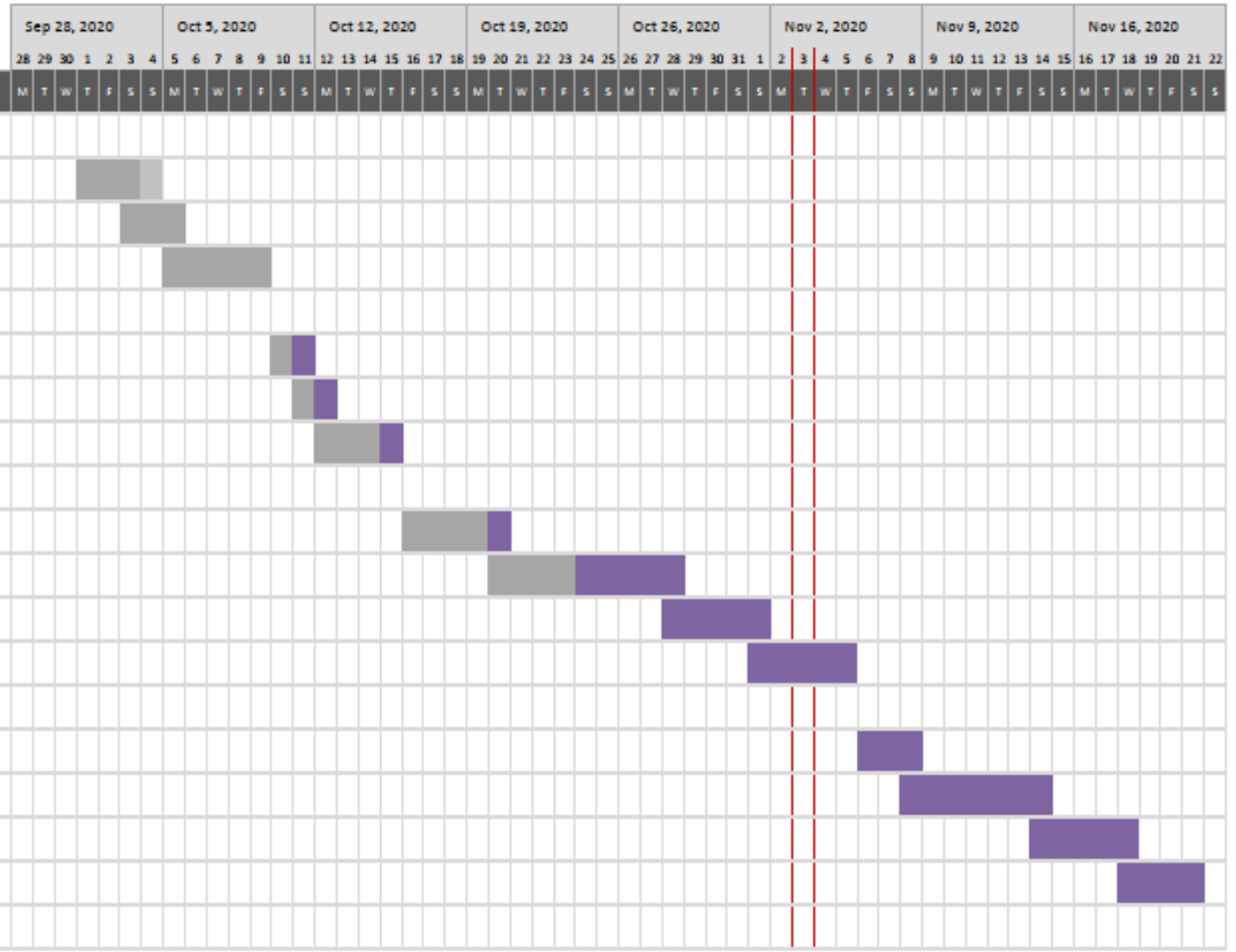
Project Plan: Gantt Chart

Predicting Disease From symptoms

Thu, 10-1-2020

1

TASK	PROGRESS	START	END
Phase 1: Problem Definition			
Problem Structuring	100%	10-1-20	10-3-20
Market survey (primary research)	100%	10-3-20	10-5-20
Business Case Development	100%	10-5-20	10-9-20
Phase 2: Technology landscape assessment			
Literature review	50%	10-10-20	10-11-20
Open libraries and patents	80%	10-11-20	10-12-20
Data collection and processing	80%	10-12-20	10-15-20
Coceptual Design			
ML Model selection	80%	10-16-20	10-20-20
Implemetation plan(coding)	50%	10-20-20	10-28-20
Performance analysis		10-28-20	11-1-20
Test plan		11-1-20	11-5-20
Phase 4: Testing and deployment			
Unit Testing		11-6-20	11-8-20
Website development		11-8-20	11-14-20
Deployment of ML model		11-14-20	11-18-20
Monitoring and maintenance		11-18-20	11-21-20



Project Plan: RASIC Chart

Project task or deliverable	Project Manager	Strategist	Designer	Front End developer	Back End developer
Problem Structuring	R	R	A	I	C
Market survey (primary research)	R	A	I	I	I
Business Case Development	R	A	C	I	I
Phase 2: Technology landscape assessment					
Literature review	C	C	A	R	R
Open libraries and patents	I	I	C	R	R
Data collection and processing	C	I	C	A	R
Coceptual Design					
ML Model selection	C	I	R	A	R
Implemetation plan(coding)	C	I	C	A	R
Performance analysis	C	I	R	A	R
Test plan	A	I	R	C	R
Phase 4: Testing and deployment					
Unit Testing	C	C	A	C	R
Website development	C	I	A	R	R
Deployment of ML model	C	C	A	R	R
Monitoring and maintenance	I	I	A	R	R

- **Responsible(R)**: The person who is ultimately responsible for delivering the project and/or task successfully.
- **Accountable(A)**: The person who has ultimate accountability and approval authority; they review and assure quality and are the person to whom “R” is accountable.
- **Supporting(S)**: The team or person(s) supporting the “real” work with resources, time or other material benefit. They are committed to its completion.
- **Informed(I)**: Those who provide input and must be informed of results or actions taken but are not involved in final decision-making .
- **Consulted(C)**: Those who provide valuable input into product design or establish quality review criteria. Their buy-in is important for successful implementation.