

# Medical Data Benchmarking for Recommender Systems



Discordant Chronic Conditions Care Challenge

Presented by: Rutuja Nimbalkar

Advisors: Dr. Tam Nguyen

Dr. Tom Ongwere

# DCC's- What are Discordant Chronic Conditions?



- Discordant Chronic Comorbidities (DCCs) are health conditions in which patients have multiple, often unrelated, chronic illnesses that may need to be addressed concurrently but may also be associated with conflicting treatment instructions.

# Major DCC's Care and Support Challenge



- Patients with DCCs often experience multiple obstacles when prioritizing treatment plans and prescriptions.
- Very few studies have addressed these complex needs to the best of our knowledge.

# Solution to DCCs Care Challenge



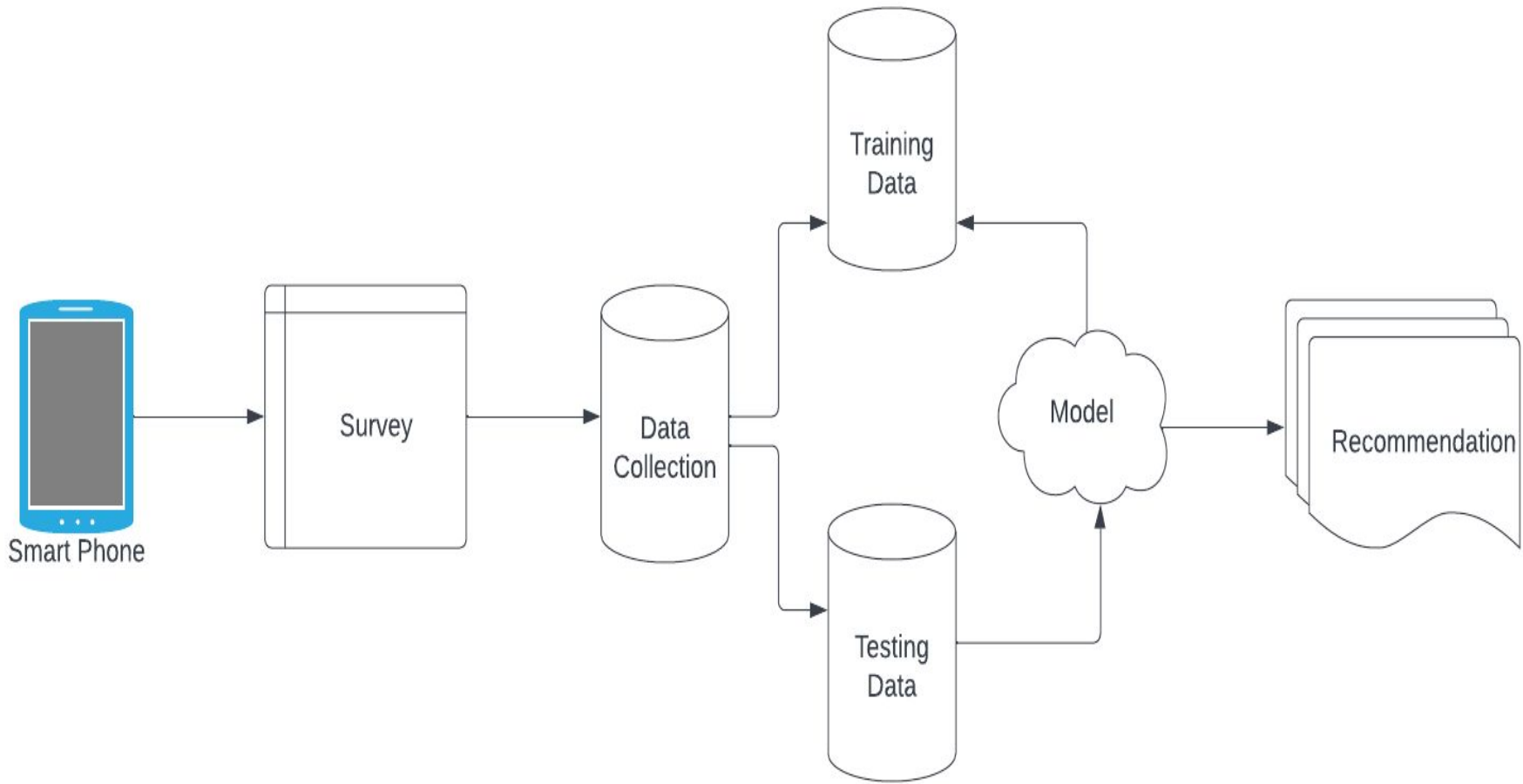
- Various machine learning (ML) or deep learning (DL) algorithms can provide these treatment recommendations and also identify drug interactions and/or conflicts.
- However currently they focussed only for patients with single disease.

# Solution to DCCs Care Challenge



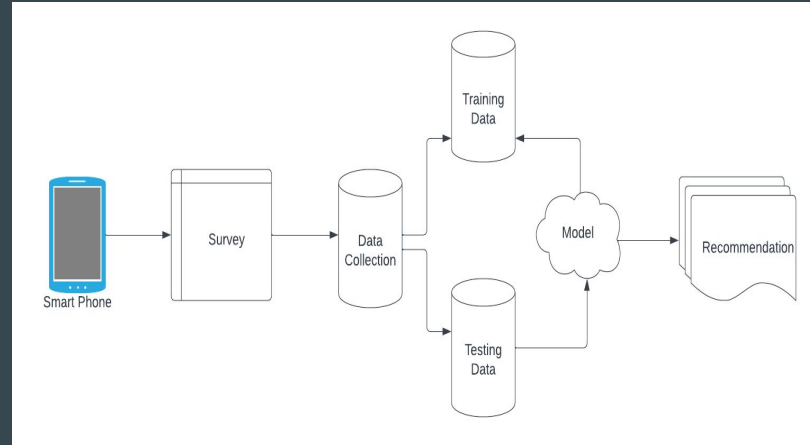
- Such machine algorithms include; support vector machines (SVM), logistic regression (LR), k-means clustering are commonly used in classification, and diagnosis of Chronic diseases and are expected to become more important in medical practice in the near future.

# Proposed System



# Proposed System

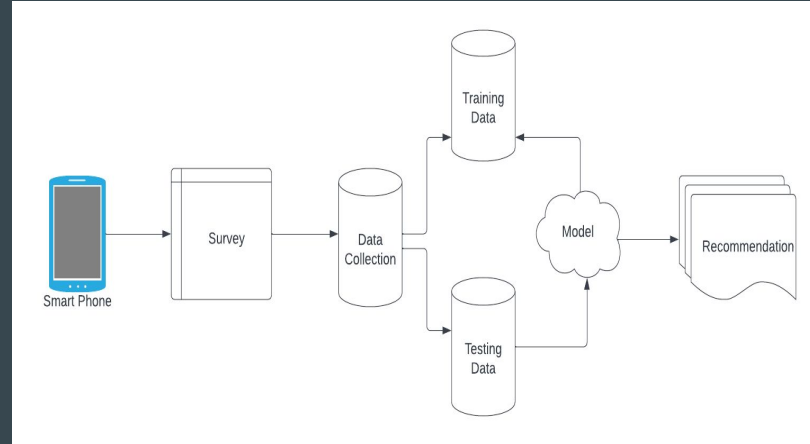
- Our proposed system consists of a survey that is to be filled by patients on the mobile devices/ app.
- This survey will collect the information regarding the concerns that are important for the patients for a particular treatment.
- Depending on the selected preferences data will be collected in an excel format and will be divided into training and testing data.





# Proposed System(continued..)

- Training data is trained on the different machine learning algorithms / models and testing data is passed into these models to record the testing accuracy of each model.
- The model with the highest accuracy is selected and the predictions from this model are the medication/Treatments consulted to the patients based on the concerns selected for a particular condition.



# Data Collection



# Dataset

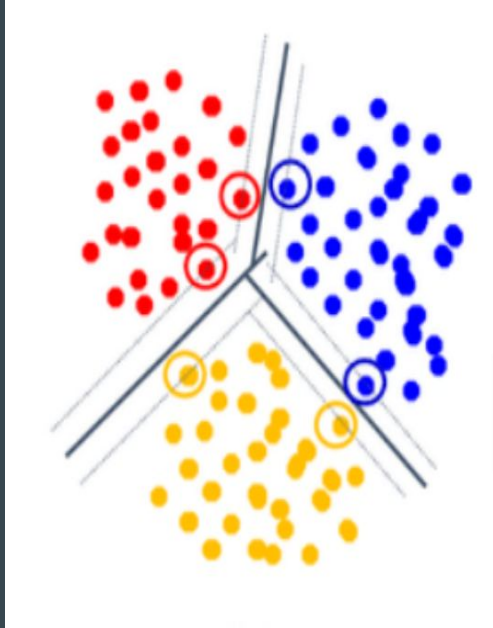
- Data was collected for the condition Type-2 Diabetes from the patients/ healthcare professionals.
- It consists of 6 features and 1 class column.
- Total samples collected = 74,  
Training data=54 and Testing data= 20

Sleep	Cost	Weight	Symptoms	Addictiveness	Effects on Sexual Performance	Class
Yes	Yes					Metformin
Yes		Yes				Empagliflozin
Yes			Yes			Empagliflozin
Yes				Yes		Empagliflozin
Yes					Yes	Metformin
	Yes	Yes				Metformin
	Yes		Yes			Metformin
	Yes				Yes	Metformin
		Yes	Yes			Empagliflozin
		Yes		Yes		Metformin
		Yes		Yes		Victoza
		Yes			Yes	Metformin
		Yes			Yes	Victoza
			Yes	Yes		Empagliflozin
			Yes		Yes	Metformin
				Yes	Yes	Metformin
				Yes	Yes	Victoza
Yes	Yes		Yes			Metformin
Yes	Yes			Yes		Metformin
Yes	Yes				Yes	Metformin
Yes		Yes	Yes			Empagliflozin
Yes		Yes			Yes	Metformin
Yes			Yes	Yes		Empagliflozin
Yes				Yes	Yes	Metformin
	Yes	Yes	Yes			Metformin
	Yes	Yes			Yes	Metformin
	Yes		Yes	Yes		Metformin
	Yes		Yes		Yes	Metformin

**Implementation of Dataset  
on  
different algorithms**

# SVM Algorithm

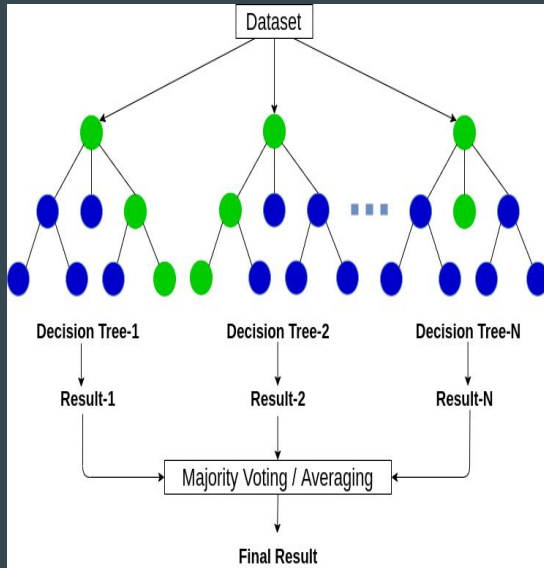
- SVM is mostly used in classification problems.



- Each data item is plot as a point in  $n$ -dimensional space (where  $n$  is a number of features) with the value of each feature being the value of a particular coordinate.
- Then, classified by finding the hyperplane that differentiates the three classes.

# Random Forest Algorithm

- Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble.



- Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.
- A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.

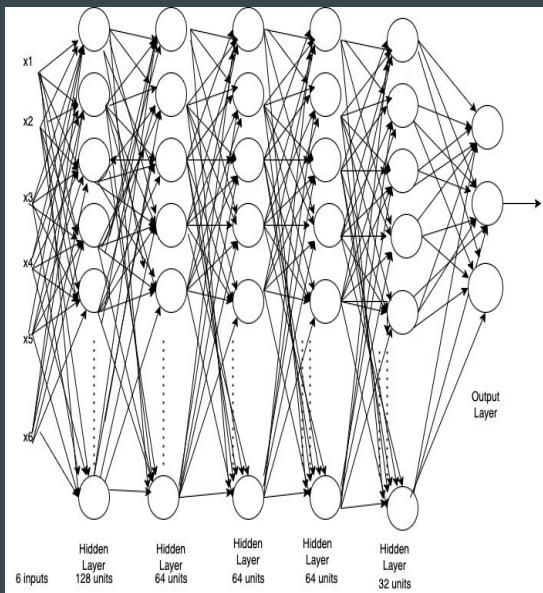
# Steps involved in Random Forest Algorithm

- Step-1 – We first make subsets of our original data. We will do row sampling and feature sampling that means we'll select rows and columns with replacement and create subsets of the training dataset.
- Step- 2 – We create an individual decision tree for each subset we take
- Step-3 – Each decision tree will give an output
- Step 4 – Final output is considered based on Majority Voting if it's a classification problem .



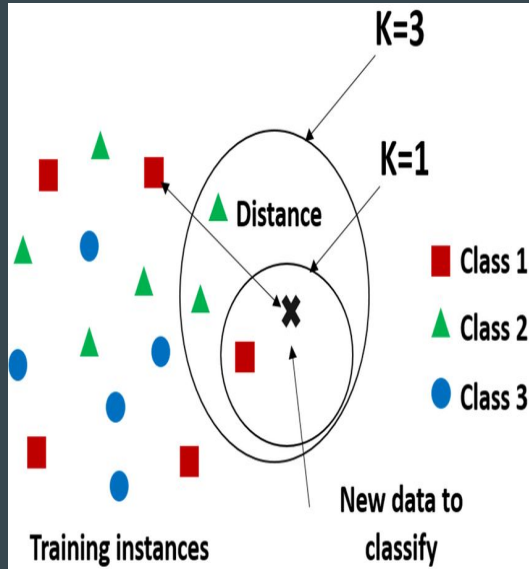
# Neural Networks

- A neural network works similarly to the human brain's neural network.



- In the supervised training, a network processes the inputs and compares its actual outputs against the expected outputs.
- This process is repeated until the errors are minimized; it means that the same set of data is processed many times as the weights between the layers of the network are refined during the training of the network.

# Nearest Neighbor Algorithm



- A supervised learning technique that computes distance between the test data and all of the training points to predict the proper class for the test data.
- The phases in the K-NN classification process are as follows:
- Choose the K value. Calculate the distance between all the training points and new data points.
- Choose the first minimum distances from the sorted list and classify the test data as a part of that training point.

# Results

# Accuracy



- Implemented our dataset on four machine learning models- SVM, Random Forest , Nearest Neighbor and Neural Networks.
- Below are the testing accuracies for these algorithms:
- SVM: 95%
- Random Forest: 85%
- Neural Networks: 57%
- Nearest Neighbor(Euclidean distance): 60%
- Nearest Neighbor(Manhattan distance): 60%

# Results



- We investigate different machine learning algorithms like Random Forest, SVM, K-NN and Neural networks on the above dataset. Following the benchmarking, SVM achieves the highest performance with an accuracy of 95%.
- In the future, we aim to collect more data for different conditions like Depression, Arthritis and evaluate different machine learning algorithms to overcome the problem of drugs interaction for multiple chronic diseases.

# Q & A's





**THANK**

**YOU!**