SLIDE 01

Good morning everyone,

We are team INFINITY from MGM University, Ch. Sambhajinagar. For this event, we will be developing an AI powered health insurance assistant; named InsureSmart. ; based on PS 07

SLIDE 02  
  
It will be a web-based platform having three major functionalities;

1. The 1st one is Evaluating the User's Health Status:
   1. This will be done by extracting pre-existing disease’s name, if any, from a diagnosis report uploaded by user.
   2. Also, underlying health condition will be identified by analyzing user’s medical test results. The test result values can be entered manually or by uploading a report. We will implement ML model trained on dataset from Kaggle. Dataset has values for 9 medical tests as its attributes and 5 labels/classes of health condition in target attribute. Link below
2. thereafter, the user will be given a questionnaire regarding his/her lifestyle choices. For eg. Smoking, drinking, exercise frequency, mental stress, occupational hazard. These answers will be taken as String (never, sometimes, frequently) stored in the form of ordinal categorical values. This will be used for calculating health score using weighted average method. The score would range from 0 (Worst) to 5(Best), The weightage for every factor would be specified by each registered insurer and the platform would take average of their inputs.
3. Next, insurers can load their plan on our platform by specifying its features like(disease it covers, health score required, monthly premium amount, addons etc.)

Thereon, Insurance plans will be recommended based upon the user’s health and lifestyle along with any addons required by him. The plans would be ranked based on degree to which they suit the user.

Also user’s will be advised regarding how of sum assured value should they apply for wrt their annual income and already existing policies’ value. This will reduce the possibilities of user’s policy application getting rejected.

Thereafter, user can contact the insurer or make direct purchases.

SLIDE 03.

Coming to the technologies that we would be implementing on our platform;

The frontend will be built using Streamlit framework,. This also allows for rapid deployment of the project on streamlit cloud. backend connections would be Python scripts. data storage would be managed by Firebase.

Text from uploaded files will be extracted using Tesseract. It is an open-source engine developed by Google. It will be complemented by PIL and pdf2image Python libraries for reading image and PDF files respectively.

The ML model to be used to identify user’s health condition can be based on either random forest Algo or Multi Layer Perceptron Algo. These 2 are best suited because the dataset is structured and its attribtutes have a non linear relationship between them. We have used a dataset from Kaggle, it’s reference link present here.

We would be implementing MLP algo on our platform. This is because, though it requires more computational power, the dataset is comparatively smaller. Also the MLP architecture can be defined as per my requirements; like setting the number of hidden layers, no. of nodes in them, which activation function to be used, no. of epochs or iterations for training the model, Later; it will be deployed as a pickle file (.pkl).

Additionally, Google Maps API can be used for displaying insurer’s address.

SLIDE 04

In future, our platform can be enhanced with additional features like:-

advanced AI insights(example – estimating Probability of claim Approval, cost-to-benefit analysis).

It may have multiple revenue streams (example - partnerships with insurance companies, API-based licensing for business-to-business collaborations (Hospitals, fintech), premium surcharges for additional insights like the ones I mentioned before.

As and when our system is optimized using huge volume of data. It will reduce the efforts required by user’s side to searching and give best suited plans

CHALLENGES – by self as better explanation in ppt.

Slide 05: -

InsureSmart would stand out from existing platforms due to its

**1.**   
Our platform streamlines the complex insurance selection process.  
Users no longer need to browse multiple websites or talk to various agents.  
With intuitive inputs and smart matching, the best policies are shown instantly.  
This enhances convenience and improves decision-making confidence.

**2.   
data-driven approach** providing accurate insights.  
Users wont have to struggle comparing multiple policies, as the system would analyze their requirement and suggests the most suitable plans.   
As and when our system is optimized using huge volume of data. It will reduce the efforts required by user’s side to searching and give best suited plans

3.   
Insurers can reach untapped audiences through digital channels.  
They gain access to user needs and preferences in real-time.  
This creates an opportunity to tailor products more effectively.  
Ultimately, it enables penetration into previously underserved markets

4..   
Users get unbiased suggestions without pressure from agents.  
Automation reduces overheads for insurers and increases transparency.  
It empowers users to make decisions on their own terms.  
We can also generate several revenue streams in form of Affiliation with insurers, API licensing for B2B partners

Slide 07.

For the development of InsureSmart, we would be referring to the following sources……….

**Additionally prepare for -**   
what is Random forest , MLP , Tesseract ,

MLP – Activation Function = mathematical function applied to the output of a neuron. ;

Max\_iter = 200; this means number of iterations/epochs to be done for training model. Each iteration has a forword propagation of data, and backward propagation for error correction. 🔹 If the model converges before 200 iterations, training stops early.  
🔹 If 200 iterations are reached and the model hasn’t converged, training stops forcefully.

Tesseract - software tool, who code is available to public, anyone can see, modify etc…

non linear relationship – though the test like Hbaic and Blood glucose are related, don’t have a simple linear relationship.

They are not simply proportional. Threshold-based health conditions ;