Product Name: Molecule
Team Name: Checkmate
Team Members: Azan Bin Zahid, Dilkasha Tarannum, Taimoor Ali

—---------- Document Details ----------

Problem Description:

As evident from various reports that Air quality in Lahore has been decreasing drastically, we have also chosen the Air Hack. Our problem statement corresponds to the fact that:

Which factors are the major contributors and how can we devise smart solutions to control those individual factors in order to improve our Air Quality Index for future by suggesting precautionary measures implemented immediately at area level?

User research:
Our users primarily consist of the Policy makers, organizations directly contributing to the Air pollutants and Government departments focusing on launching new products in order to improve the Air quality in coming years.

Dimensions:
There could be many possible reasons for the depreciation of Air Quality Index but our main focus will be driven towards 4 dimensions which are the primary contributors to the Air pollution in Lahore City. These dimensions include:

1. Traffic Congestion
   High private transport vehicles and dense traffic areas are two main causes of traffic congestion in Lahore. In areas like Anarkali, Old Lahore we have more density of traffic, whereas in comparatively developed areas like Johar Town, Model Town we have high private transports which contribute to the increased traffic congestion and ultimately harm the AIr Quality Index in Lahore.
2. Population Inflation
   We have witnessed an increase in population of Pakistan but specifically Lahore has been the host to many citizens migrating from other cities which contributes to increased resource to people ratio, this contributes to the Air pollution since the place has been overcrowded.
3. Deforestation
   Deforestation is contributing to Air pollution since the amount of Oxygen required is not being fulfilled by the current forest count and the Carbon oxides are being emitted in the environment very rapidly which would have been compensated by the CO2 consumption of Trees. Likewise burning of wood, combustion of fossil fuels are increasing the percentage of Carbon, Sulphur and Nitrogen Oxides which are not being absorbed or processed by trees and hence end up in the form of Acid rain, Smog and Air pollution.
4. Industrial Wastes

Increased large scale industries and improper waste management of these industries is another leading cause of harmful oxides emission which goes into air without being treated properly which adds to the Air pollution.

Data Sources:
Our main focus is to perform Data Analysis and predict future hazards in order to take timely precautions and work for the improvement of the Air Quality Index. We will gather data from different sources and perform ETL (Extract, Transform and Lead) Techniques to integrate data coming from different sources and transform all data sets to a unified data format that will be processed by our system. Data Analysis will be further performed on unified data in order to train our models and find the optimal correlation equations that have the ability to predict future instances with great accuracy and minimalistic errors. Our major focus will be driven towards 4 dimensions as mentioned above and their data sources will be as follows:

1. Traffic Congestion
   As we are talking about the smart solutions, we will improvise the traffic congestion data from Google maps instead of collecting data, manipulating and analysing it from scratch. Google maps use the Mobile signals coming back due to the location features enabled on smart phones and based on those signals we receive the traffic congestion heatmaps along with the suggestion on different routes. Apart from that we can also take data from Online sources and analyse these data points.
2. Population Inflation
   Population Inflation data will be taken from the online platform where Population Census is being recorded but not used in this aspect.
3. Deforestation
   We are taking the deforestation information from our past project "Forest Management Portal" which has a follow up published research paper. This project was taking satellite images, performing image processing and calculating the vegetation indices based on the spectral bands information which provided us with information regarding the low vegetation areas, high vegetation areas and the records of the endangered tree species so that we can focus on preserving those tree species and plant more trees in the areas having high rate of deforestation. So our assumption is based on having deforestation data from there and from online resources. This data will be used to analyse the correlation between the Forests and Air Purification.

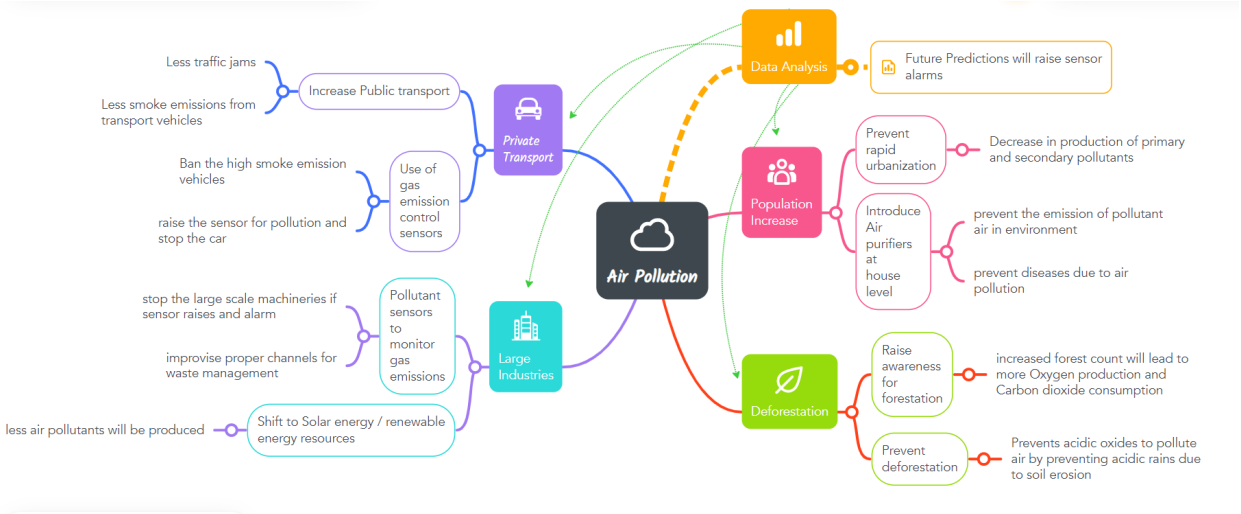
Assumptions:
Due to time constraint we could not gather runtime data so we have our analysis based on limited data. However, the predictive models have been trained using the retrospective available data.

Methodology and Processes:
WIth the above mentioned Data acquisition assumptions, we proceed to our main product functionality. We are focusing on the SCAMPER approach delivering the Smart solutions for AIr pollution which could be implemented immediately and can have a clear impact in the near future.
We will be performing following techniques on our data to make it an easy to use solution:

1. Data Engineering: ETL techniques are used in order to gather our data from different parameters in a unified data format
2. Machine Learning: We are using the simplistic machine learning model, Linear Regression to train our models based on the retrospective data and use these models to predict the future.
3. Big Data Analysis: We are using the 3D Network Graphs mapped on the series of time events. These graphs will represent the weighted impact of all four dimensions that we have diagnosed along with their power coefficient representing the intensity of these individual dimensions.
4. Software Engineering: We have gone through the whole process of Web development; starting from the User research, Identifying the problem statement, Ideation, Mind Mapping, Storyboards, User cases, Low-Fii and High-Fi Prototype
5. Product Design: We have specially focused on SCAMPER technique, making a Minimal Viable Product (MVP) in order to follow the process of substituting solutions, Combining them, making them adaptable, leaving a room for making modifications, elimination and coding in a scalable fashion so that product could be reverse engineering in another series of Product Development Life Cycle.
6. Human Computer Interaction: Our colour palette is in accordance with the Norman's design principles. We have tried to implement the visibility, responsiveness, feedback, mapping and consistency in our product.

Technologies Used:
Frontend Technologies:
Backend TEchnologies:
Data Engineering: We are using Python for performing the ETL techniques on raw data and transforming it in unified data format.

Mind Map:



User Stories:

User :
Policy Maker

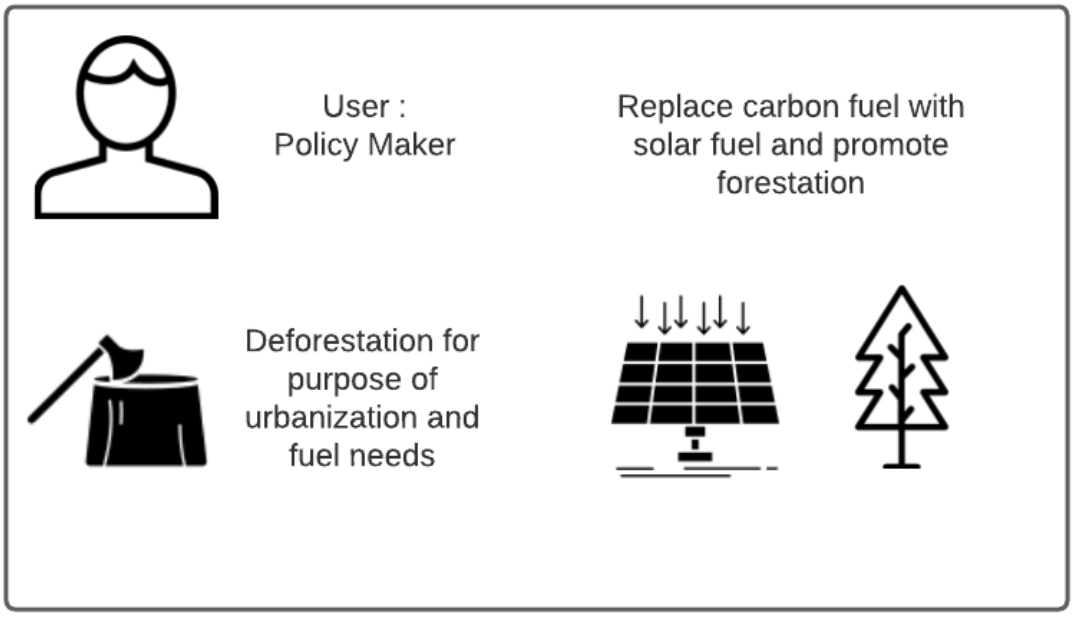
Future predictions will raise alarms and suggest you to install filters

Industrial Pollution is leading to high Air Pollution, how should we contriol it ?

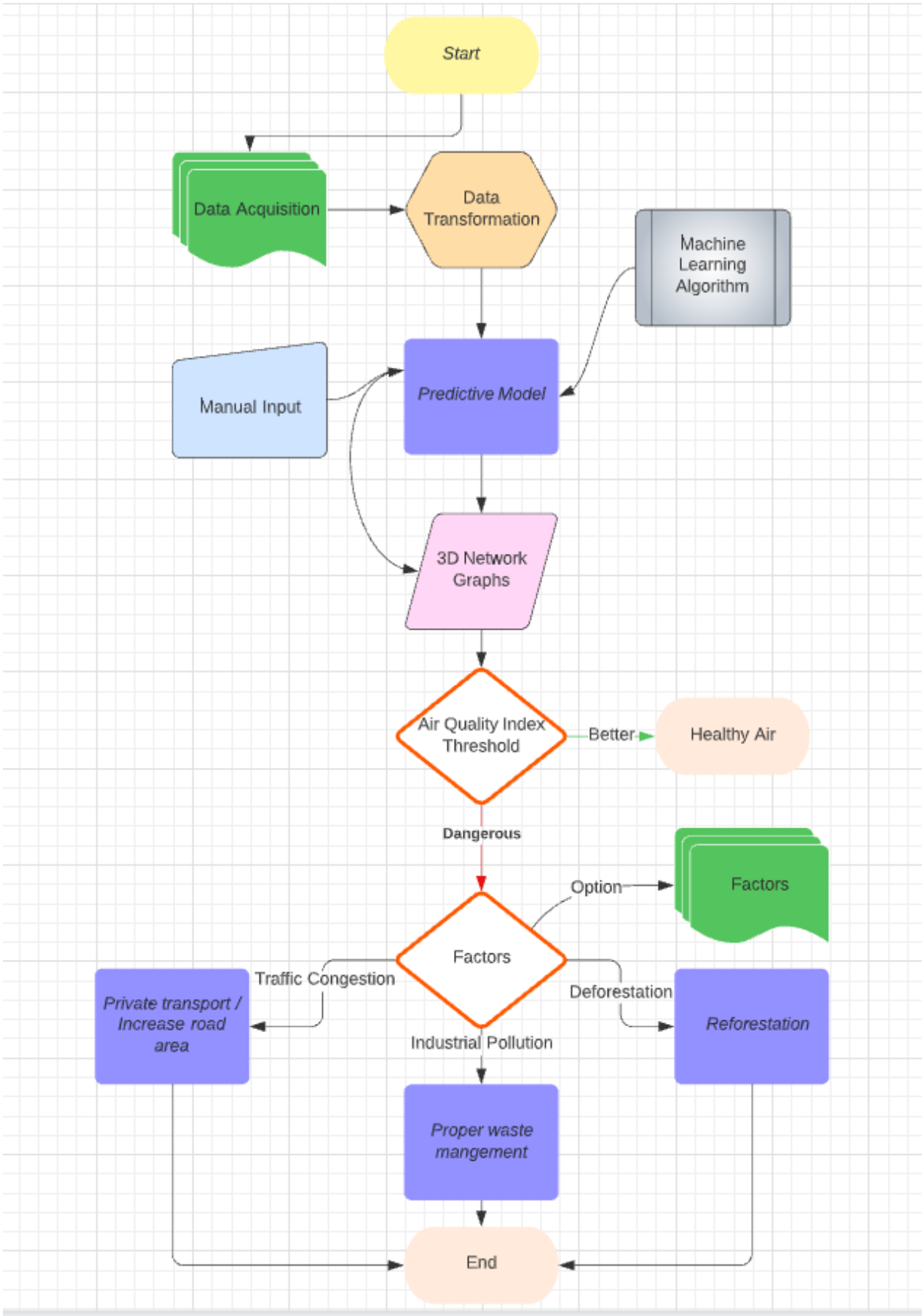
User :
Government Organizations

Future predictions will raise alarms and suggest you to use public transport

Traffic jams and Trafic smoke is contributing to high index air pollution


User :
Policy Maker
Replace carbon fuel with solar fuel and promote forestation
Deforestation for purpose of urbanization and fuel needs

Flow chart:


Start
Data Acquisition
Data Transformation
Machine Learning Algorithm
Manual Input
Predictive Model
3D Network Graphs
Air Quality Index Threshold
Better
Healthy Air
Dangerous
Option
Factors
Factors
Traffic Congestion
Deforestation
Private transport / Increase road area
Reforestation
Industrial Pollution
Proper waste mangement
End

Product Growth in Future:
Our project is easily scalable, adaptable and has the best 3D Network Analysis graph to comprehend the need of hour and implement the smart solutions immediately after the diagnosis at local level.