**DEVELOPMENT OF A SMOKING CESSATION SUPPORT SYSTEM**

**MILESTONE: PROJECT PROPOSAL**

**GROUP 90**

Sai Srinivas Madamraju

Sri Mahalakshmi Harika Punati

Contact number

(609)-803-1239

(510)-460-8178

Email

[**madamraju.s@northeastern.edu**](mailto:madamraju.s@northeastern.edu)

**punati.s@northeastern.edu**

Percentage of effort contribution by Sai Srinivas Madamraju: 50%

Percentage of effort contribution by Sri Mahalakshmi Harika Punati: 50%

Signature of student 1: Sai Srinivas Madamraju

Signature of student 2: Sri Mahalakshmi Harika Punati

Submission Date: 02-03-2023

Problem Setting:

Smoking causes an estimated around 7 million deaths annually, which is a serious public health hazard. It is a major contributor to avoidable deaths and is linked to a variety of illnesses, such as heart disease, chronic obstructive pulmonary disease (COPD), and lung cancer. Additionally, smoking not only harms the smoker, but it also affects people nearby who are exposed to second-hand smoke. The financial burden of smoking on society is also significant, with billions of dollars spent each year on healthcare costs related to smoking-related illnesses. Despite the well-known risks of smoking and the availability of several initiatives to help people quit, many people still smoke. Therefore, there is a need for effective interventions to help individuals quit smoking and prevent the initiation of smoking in young people.

Problem Definition:

Despite the availability of evidence-based treatment for smoking cessation, success rates remain low, with only a small percentage of participants achieving abstinence. It is important to educate them about the serious medical issues brought on by smoking in order to increase the likelihood of success. The model analyses the inputs and determines if the subject smokes or not. However, using each of these variables separately can provide results that are hard for patients and medical professionals to understand and use. To understand the likelihood that each smoker would successfully quit, a prediction model employing machine learning techniques may be more efficient.

Data Source:

**Development of a smoking cessation support system data set has been taken from smoker status prediction dataset using Kaggle datasets (**[**https://www.kaggle.com/datasets/gauravduttakiit/smoker-status-prediction**](https://www.kaggle.com/datasets/gauravduttakiit/smoker-status-prediction)**). The data description is from the above mentioned source.**

Data Description:

The dataset consists of 38,985 rows and 23 columns. It contains 3 categorical variables and 20 numerical variables. The columns are comprised with data like age : 5-years gap, height(cm), weight(kg), waist(cm) : Waist circumference length, eyesight(left), eyesight(right), hearing(left), hearing(right), systolic : Blood pressure, relaxation : Blood pressure, fasting blood sugar, Cholesterol : total, triglyceride, HDL : cholesterol type, LDL : cholesterol type, haemoglobin, Urine protein, serum creatinine, AST : glutamic oxaloacetic transaminase type, ALT : glutamic oxaloacetic transaminase type, Gtp : γ-GTP, dental caries, smoking