

# Public Health Awareness Campaign Analysis Project Design and Innovation

Date	04-10-2023
Team ID	716
Project Name	6112-Public Health Awareness Campaign Analysis using Data Analytics

## Table of Contents

1	Introduction
2	Problem Statement
3	Design and Innovation Strategies
3.1	Data Collection and Feature Engineering
3.2	Data Pre-processing
3.3	Model Selection and Training
3.4	Data integration with IBM Cognos
3.5	Public Health awareness report and dashboard
3.6	Deploying and Monitoring the Model
3.7	Continuous Learning
4	Conclusion

### 1. Introduction

The objective of this project is to provide an in-depth analysis of the design and innovation strategies for developing a public health awareness campaign using data analytics with cognos tool. A public health awareness campaign is a powerful tool for promoting well-being within communities and this project aims to utilize innovative approaches to enhance prediction accuracy and reliability.

### 2. Problem Statement

The tech industry has been known for its fast-paced and demanding work environment, which can often lead to increased stress and mental health challenges among its employees. It is essential to understand the extent of these issues, identify contributing factors, and explore potential solutions to improve the mental well-being of tech professionals. This survey aims to investigate the current state of mental health in the tech industry the factors affecting it, and the attitudes towards seeking help and support among tech workers.

### 3. Design and Innovation Strategies

#### 3.1. Data Collection and Feature Engineering

**Innovation:** Comprehensive Data Gathering

**Multi-Source Data Integration:** Instead of relying solely on internal data sources, consider integrating external data, such as country united states,france,canada etc.,age,treatment,benefits,consequence,seek help,mental health to understanding of mental health behaviour and external factor that may influence mental health.

**Surveys and Feedback:** Conduct surveys and collect workers feedback systematically. Analyze unstructured text feedback using natural language processing (NLP) techniques to extract valuable insights about worker sentiments and pain points.

**Sentiment Analysis:** Utilize sentiment analysis on workers feedback and reviews to generate sentiment-based features, such as sentiment scores or sentiment trend indicators.

### 3.2. Data Pre-processing

Data pre-processing is a critical step it involves cleaning and preparing the data to ensure it is suitable for analysis and modelling.

**Innovation:** Natural Language Processing (NLP) for Unstructured Data

**Outlier Detection and Handling:** Develop a custom NLP pipeline that includes tokenization, lemmatization, sentiment analysis, and named entity recognition to enhance the quality of textual data.

Utilize Natural Language Processing (NLP) techniques to pre-process textual data (e.g., property descriptions) and extract valuable information.

Develop a custom NLP pipeline that includes tokenization, lemmatization, sentiment analysis, and named entity recognition to enhance the quality of textual data.

Handle missing data with innovative methods such as K-nearest neighbours imputation tailored for spatial data, reducing information loss.

### 3.3. Model Selection and Training

**Innovation:** Ensemble Learning and Deep Learning Integration.

**Algorithm Selection:** Consider various algorithms suitable for classification tasks,

**Regression Analysis:** Linear regression or logistic regression can be used to identify relationships between mental health variables and other factors such as workload, support systems, or demographic information.

**Cluster Analysis:** Cluster analysis helps identify groups of individuals with similar mental health profiles. This can be useful for segmenting survey respondents based on their mental health needs.

**Decision Trees:** Decision tree algorithms like CART or Random Forest can be used to create predictive models that classify individuals into different mental health categories based on survey responses and other features

**Cross-Validation:** Implement k-fold cross-validation to assess model performance more robustly and reduce the risk of overfitting. Normalize or standardize features if required by the chosen algorithm.

### 3.4. Data Integration with IBM Cognos

**Data Extraction:** Use IBM Cognos Data Integration Studio or other ETL (Extract, Transform, Load) tools to extract data from your various sources. Cognos Data Integration Studio is a powerful tool that can connect to a wide range of data sources.

**Data Transformation:** After extracting the data, perform necessary transformations to ensure that it is in the right format for analysis. This may include cleaning, filtering, aggregating, and merging data from different sources.

**Data Modelling:** Design a data model that represents the integrated dataset. This model should define how different data elements relate to each other, creating a coherent structure for analysis.

**Data Exploration and Analysis:** Use IBM Cognos Analytics to explore and analyse your integrated data. You can create reports, dashboards, and visualizations to gain insights into patterns and potential predictors.

### 3.5. Public Health awareness report and dashboard

Public health reports and dashboards is essential for visualizing and communicating the results of your public health awareness models effectively. These reports and dashboards provide actionable insights to help and make decisions.

**Innovation:** Sentiment analysis of public health awareness

This involves using natural language processing to determine the sentiment expressed in survey responses, helping to gauge the emotional well-being of respondents.

### 3.6. Deploying and Monitoring the Model

**Innovation:** Model Interpretability

**Model Integration:** Integrate your trained churn prediction model into your production environment. This typically involves deploying the model as an API or a service that can receive real-time data.

**Data Quality Monitoring:** Continuously monitor the quality of input data to ensure it meets the model's requirements. Data quality issues can lead to inaccurate predictions.

### 3.7. Continuous Learning

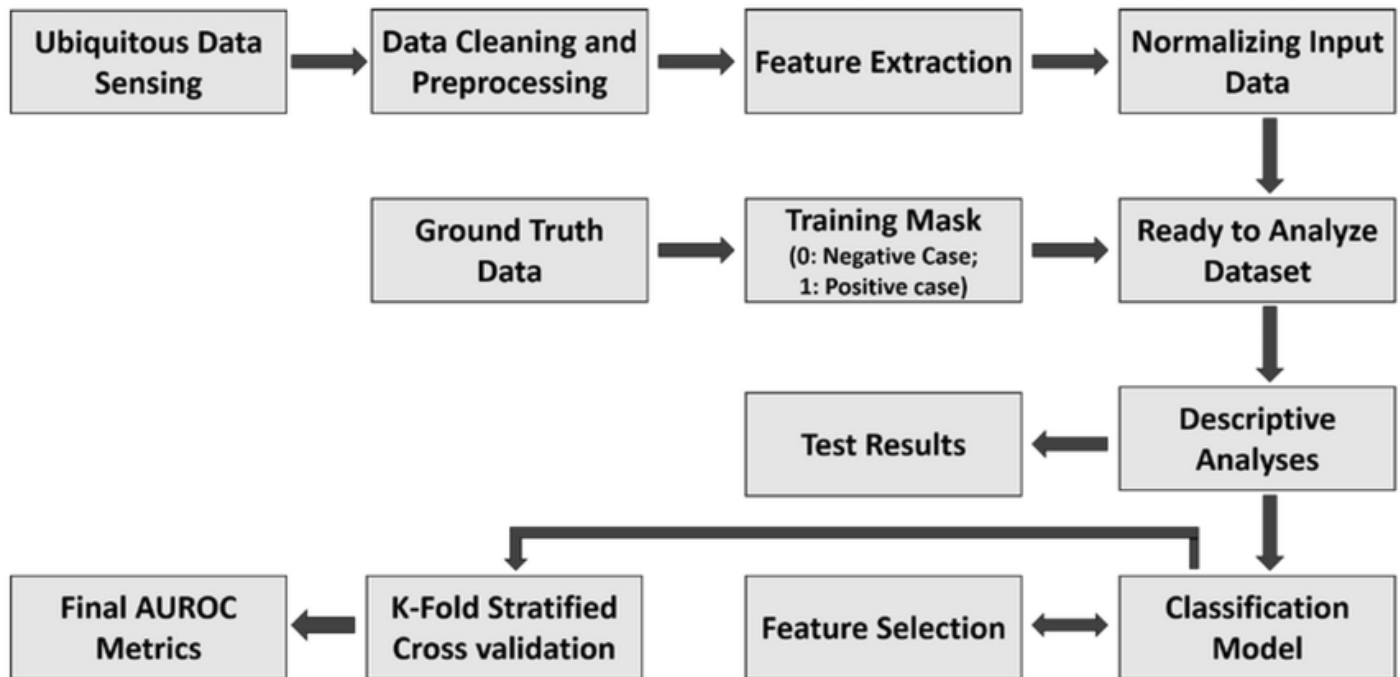
**Innovation:** Model Maintenance and Improvement

Establish a continuous learning framework that incorporates user feedback and new data to update and enhance the model's performance.

Regularly retrain the model to adapt to changing market dynamics and ensure long-term accuracy.

Note: In the diagram below, we've depicted the key components and interactions described in sections 3.1 to 3.7, offering a clear and concise overview of our solution architecture. This

visualization simplifies the complex concepts and relationships discussed in those sections, making it easier for the reader to grasp the overall design and innovation strategies at a glance.



#### 4. Conclusion

The public health awareness project employs a holistic approach to address the challenges of health and awareness accurately. By integrating innovative strategies such as comprehensive data collection, NLP for unstructured data, ensemble learning, geographic analysis, market sentiment analysis, and continuous learning, this project aims to develop a robust and reliable model. Through a combination of cutting-edge technologies and techniques, we aspire to provide a comprehensive and insightful solution for the public health awareness system.