

FITNESS CONSUMER ANALYSIS

TABLE OF CONTENT

01

Project Overview

Key Attributes: Age, Gender, Education, Occupation, Exercise Frequency, Wearable Duration, Wearable Frequency, Track Data Frequency, Routine Impact, Motivation Impact, Enjoyment Impact, Engagement, Community Connection, Goal Impact, Health Impact, Sleep Impact, Wellbeing Impact, Decision to Exercise More, Decision to Buy Products, Decision to Join Gym, Decision to Change Diet

02

Libraries and Data Handling

Numerical Operations and Data Manipulation: NumPy and Pandas

Data Visualization: Matplotlib and Seaborn

Statistical Modeling and Analysis: Statsmodels, Scipy, Scipy Chi-square Test, Scipy One-way ANOVA Test, Ordinary Least Squares Regression

Model Selection and Evaluation: Train-Test Split, Logistic Regression, Naive Bayes Classifier, K-Nearest Neighbors Classifier, Decision Tree Classifier, Support Vector Classifier, Random Forest Classifier, Gradient Boosting Classifier, Permutation Importance, Evaluation Metrics

Data Processing: Label Encoder, OneHotEncoder, and Standard Scaler

Data Formatting: Pprint

03

Visual Insights

Distribution of User Attributes: The representation of various user characteristics and responses using bar graphs as a visualization

04

Key Findings and Business Impact

Cramer's V: A measure of association between two nominal variables, indicating the strength and significance of their relationship, ranging from 0 (no association) to 1 (perfect association).

FITNESS CONSUMER ANALYSIS

TABLE OF CONTENT

05

Data Analysis Techniques

Descriptive Statistics: A set of statistical tools used to describe and summarize the main features of a dataset.

ANOVA (Analysis of Variance): A statistical method used to compare the means of three or more groups to determine if there are statistically significant differences among them.

06

Implementation of ML

Prediction Models: Decision to Exercise More, Decision to Buy Products, Decision to Join Gym, Decision to Change Diet

Machine Learning Models: Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine, Gaussian Naive Bayes, K-Nearest Neighbor, and Decision Tree

Evaluation Metrics: Accuracy, Precision, Recall, F1

07

Advanced Analysis and Conclusion

Feature Importance Analysis: Highlights the most influential factors that affect each model.

Conclusion: Summary of insights derived, implications for future strategic decisions.

08

Appendix

Code Snippets: https://github.com/nisodump/csst-104-finals/blob/main/04_Code_Snippet_ML.pdf

Datasets: https://github.com/nisodump/csst-104-finals/blob/main/05_Sample_Dataset_Content.csv

Google Colab Link: https://colab.research.google.com/drive/1GvWsWECFTKWwOTcXmrHUKFa_O0XFzL4V?usp=sharing