

ArindamChatterjee_DataMiningProject

Arindam Chatterjee, Avijit Mallik, Takehiro Hasimoto

2023-09-02

R Markdown Document

Title: Visual Nudge To Inverse Monetary Preferences

Authors: Arindam Chatterjee, Avijit Mallik, Takehiro Hasimoto

Abstract Diabetes is a prevalent chronic disease in the US, affecting millions of people and placing a significant financial burden on the economy. So in this paper, to contribute decreasing diabetes, we executed four analysis on four important pillars. Those are (1) to build the best predict diabetes model, (2) to construct the risk score for diabetes, (3) To identify the critical factors causing Diabetes, using correlation matrix, PCA analysis & variable importance plot (using a random forest model), and (4) find other important factors that we have not yet got from academic paper. For the first analysis, we got the linear probability model with 92 variables (21 variables and 71 cross term variables). For the second analysis, we could get the crucial variables to increase the risk score of diabetes is High blood pressure and General health and BMI. While the variable that is not related to the diabetes, i.e. it is that variable that makes the risk score to decrease is 'Heavy Alcohol Consumption.' From the third analysis, we found from the correlation matrix which shows that there is a high positive correlation between Diabetes & High BP, High cholesterol, age, Heart disease, BMI, and also general health..

[Click here to view the PDF](#)