Classification of Patients with Abnormal Blood Pressure

Background

Hypertension with aging is a major medical concern even in this ultramodern era of advanced therapies. Preliminary testing is a key element in analyzing the procedures of mild abnormality of blood pressure for a substantial period of time, but presently has a limited value in the prediction of progression to hypertension. Although **hypotension**, low blood pressure in common parlance, is less common among the ones with abnormal blood pressure, but the adverse effect it has on their health is equivalent to that of hypertension.

Although recent studies have hypothesized that, keeping other factors constant, hemoglobin level is positively associated with blood pressure in a large cohort of healthy individuals, but genetic causes are also prominent in some of individuals. However, there are numerous other factors which determine whether an individual is likely to develop this abnormality.

Data scientists lay out the hypothesis that newer statistical classification methods derived from data mining and machine learning methods are capable of reducing the prediction error manifolds and help cardiologists to conduct a two-tailed preliminary prediction of abnormality of blood pressure in an individual.

Objective

Employing statistical techniques, conduct a preliminary prognosis of Hypertension/hypotension, based on the level of **hemoglobin** and **genetic history** of the individual.
[Please refer to the Data Dictionary (on the next page) to know more about these variables]

Deliverables

(i) Lay out an **approach plan**, consisting of:
   a. Your **understanding of data**, based on a preliminary exploratory analysis
   b. **Different** traditional as well as state-of-the-art statistical **techniques**, which you are going to use to come up with different models to meet the objective

(ii) Contrast the **pros and cons of** applying each **technique** on this problem

(iii) **Build a model** using the most promising technique on the dataset.
Model Validation is supposed to be done on the test dataset *(to be given to you during your case-study presentation)*

(iv) What would be your **approach**, **if** there were **other variables also** in the data: *Smoking, obesity (BMI), Lack of physical activity, salt content in the diet, alcohol consumption per day, Level of Stress, Age, Sex, Pregnancy, Chronic kidney disease* and *Adrenal & thyroid disorders*.

[Please refer to the Data Dictionary (on the next page) to know more about these variables]

Share the software code used to execute each technique or operation.
The approach plan, model outputs, model diagnostics, software codes etc. are supposed to be shared in an Excel file.
Support your deliverables with exhibits and slabs, wherever required.

Note: The data are hypothetical.

Data Dictionary

Total Number of Patients (N) = 2000

Variable	Position	Variable Label	Value Labels	Measurement Level	Role
Patient_Number	1	Patient Number	Not Applicable	Ratio	None
Blood_Pressure_Abnormality	2	Blood Pressure Abnormality	0 = Normal 1 = Abnormal	Nominal	Target
Level_of_Hemoglobin	3	Level of Hemoglobin (g/dl)	Not Applicable	Ratio	Input
Genetic_Pedigree_Coefficient	4	Genetic Pedigree Coefficient*	Not Applicable	Ratio	Input
Age	5	Age	Not Applicable	Ratio	Input
BMI	6	BMI	Not Applicable	Ratio	Input
Sex	7	Sex	0 = Male 1 = Female	Nominal	Input
Pregnancy	8	Pregnancy	0 = No 1 = Yes	Nominal	Input
Smoking	9	Smoking	0 = No 1 = Yes	Nominal	Input
Physical_activity	10	Physical activity (No. of steps/day)	Not Applicable	Ratio	Input
salt_content_in_the_diet	11	Salt content in the diet (mg/per day)	Not Applicable	Ratio	Input
alcohol_consumption_per_day	12	Alcohol consumption per day (ml/day)	Not Applicable	Ratio	Input
Level_of_Stress	13	Level of Stress (Cortisol Secretion)	1 = Less 2 = Normal 3 = High	Ordinal	Input
Chronic_kidney_disease	14	Chronic kidney disease	0 = No 1 = Yes	Nominal	Input
Adrenal_and_thyroid_disorders	15	Adrenal and thyroid disorders	0 = No 1 = Yes	Nominal	Input

******Genetic Pedigree Coefficient*** *(GPC) of an individual for a particular disease is a continuum between 0 and 1, where*
GPC ***closer to 0*** *indicates very* ***distant occurrence*** *of that disease in her/his pedigree, and*
GPC ***closer to 1*** *indicates very* ***immediate occurrence*** *of that disease in her/his pedigree]*