Knowledge Transfer associate in Machine Learning
Technical task

Goal of the task: To assess machine learning knowledge and problem-solving strategies.

Task: For the interview, prepare a presentation of maximum 30-minutes to demonstrate solution strategies to solve atleast one of the below problems using machine learning. Any details that can be included in the presentation to support your case and demonstrate the viability of the proposed solution is welcome. The presentation time of 30-minutes will include questions and technical discussions. This will be followed by a 15-minute general discussion.

Link to Dropbox folder with dataset for the below tasks: https://shorturl.at/ASTY5

NOTE: It is not expected of you to solve the problem within a week!

GUIDELINES FOR PRESENTATION

Please illustrate your approach through a presentation. Some guidance for the preparation of the presentations includes the demonstration of:

1. Your comprehension of the problem (max 2 slides)
2. Solution strategy
    a. Introducing of machine learning, you are using for a general audience (max 3 slides)
    b. Solution strategy with details of the machine learning used, for a technical audience (max 4 slides)
    c. Potential coding steps to build this (max 3 slides)
3. Important aspects to consider building the network and critical hyperparameters (max 2 slides)
4. Potential influence of hyperparameter tuning (max 2 slides)
5. What metrics will be used to test the validity of the predictions from ML (max 1 slides)
6. Usage of technologies like Streamlit to build easy to use dashboards that can be maintained with ease by a non-software engineer (max 2 slides)
7. Usage of CI/CD pipeline, comprehension of docker, git and other software pipeline technologies (max 1 slide)

PROBLEM 01 (TABULAR DATA)

General description: The data from 554 children enrolled in the Orinda Longitudinal Study of Myopia (OLSM) as nonmyopes with baseline data from the third grade were evaluated to develop a predictive profile for later onset of juvenile myopia [4]. Myopia was defined as at least -0.75 D of myopia in the vertical and horizontal meridians of the right eye as measured by cycloplegic autorefraction (n = 45 children).

Goal: Can you use the given data to provide predictions about the onset of juvenile myopia. What techniques will you use and why? Why do you think that the technique will be effective

in the prediction? What do you think could be in a software dashboard that the clinician can use to help with the prediction.

Dataset description: The dataset is provided in folder "Problem 01" in the above shared Dropbox folder. The column name and related description of the dataset, provided in the excel sheet, is provided below. A supplementary paper is also provided to help with understanding the problem.

1. STUDY YEAR: Year subject entered the study (Year)
2. MYOPIC: Myopia within the first five years of follow up (0: No, 1: Yes)
3. AGE: Age of the child:
4. GENDER: Male (0) or Female (1)
5. SPHEQ: Spherical Equivalent Refraction (diopter)
6. AL: Axial length (mm)
7. ACD: Anterior chamber depth (mm)
8. LT: Lens thickness (mm)
9. VCD: Vitreous chamber depth (mm)
10. SPORTHR: How many hours per week outside of school the child spent engaging in sports/outdoor activities (Hours per week)
11. READHR: How many hours per week outside of school the child spent reading for pleasure (Hours per week)
12. COMPHR: How many hours per week outside of school the child spent playing video/computer games or working on the computer (Hours per week)
13. STUDYHR: How many hours per week outside of school the child spent reading or studying for school assignments (Hours per week)
14. DIOPTERHR: Composite of near-work activities (Hours per week)
15. MOMMY: Was the subject's mother myopic? (0: No, 1: Yes)
16. DADDY: Was the subject's father myopic? (0: No, 1: Yes)

PROBLEM 02 (IMAGE DATA)

General description: The retinal scans or also commonly known as fundus images are used to image the eye. In the recent years, there have been several databases like FIVES [1], PAPILA [2], TREND [3] etc. that have accelerated the usage of machine learning for medical technologies. The fact that the retinal microvasculature can be imaged with reasonable ease in a non-invasive manner and without the need for dyes/drugs has made it attractive as a potential "window to health." In particular for applications to develop risk assessment for rapidly increasing myopia, retinopathy, glaucoma etc. in today's ageing population. One of the recent popular beliefs is that the decrease in retinal vascular complexity has a correlation with cognitive impairment in the aging population. The complexity of the network is often measured by fractal measurements etc.

Goal: Choose one of the three datasets (FIVES / PAPILA / TREND). Can you use the given data to provide predictions about the onset of eye diseases. What techniques will you use and why? Why do you think that the technique will be effective in the prediction? What do you think could be in a software dashboard that the clinician can use to help with the prediction.

REFERENCES

1. FIVES dataset (https://shorturl.at/AHL47)
2. PAPILA dataset (https://doi.org/10.6084/m9.figshare.14798004.v1)
3. TREND dataset (https://zenodo.org/records/4521044)
4. Zadnik, K., et.al., “Ocular predictors of the onset of juvenile myopia,” Invest Ophthalmol Vis Sci., v.40, pp. 1936-43 (1999)