# The Health Case - Canvas (v0.1) Designed for: Data Science – Post Graduate class Designed by: Prof. Edson Neto Date: 19/02/2021 Iteration: 1 .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Decisions**  How are predictions used to make decisions that provide the proposed value to the end-user?   * Correlação entre as comorbidades e o diagnóstico de derrame pode identificar aqueles fatores que mais causam a enfermidade ou quais são os maiores indicadores que alertam o surgimento desta; * As doenças relacionadas podem, através de suas premissas de cura, também contribuir para a redução de um quadro cumulativo de derrame. | **ML task**  Input, output to predict, type of problem.   * Análise de correlação entre as comorbidades; * Relação individual e em conjunto das comorbidades e o derrame; * Matrizes de dispersão para avaliação de impacto; * Probabilidade de ocorrência; * **Input:** gender,age,hypertension,heart\_disease,ever\_married,work\_type,Residence\_type,avg\_glucose\_level,bmi,smoking\_status,stroke; * **Output:** nível do derrame pelo acúmulo de comorbidades. | **Value Propositions**  What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?   * Dentre os inputs citados, qual a relação destes com o desenvolvimento do prognóstico de derrame? * Quais são os principais fatores que levam ao derrame? * Quais prevenções podem ser feitas para evitar o surgimento deste enfermidade? * Existem “conceitos populares” que envolvem a situação de derrame, mas que se provam errados pelos dados apresentados? | **Data Sources**  Which raw data sources can we use (internal and external)?   * Base de dados ‘healthcare-dataset-stroke-data’ do site Kaggle; * **Link:** ‘https://www.kaggle.com/fedesoriano/stroke-prediction-dataset?select=healthcare-dataset-stroke-data.csv’ | **Collecting Data**  How do we get new data to learn from (inputs and outputs)?   * **Input:** gender, age, hypertension, heart\_disease, ever\_married, work\_type, Residence\_type, avg\_glucose\_level, bmi, smoking\_status, stroke (todos os dados foram considerados relevantes); * **Output:** relação derrame x comorbidades; derrame x situações sociais; e relações correlacionadas entre as características de entrada. |
| **Making Predictions**  When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?   * Os primeiros dados a serem levantados serão base para o aperfeiçoamento do sistema preditivo; * A máquina será desenvolvida para aprender a detectar graus de perigo da caracterização da doença; * A medida que novos inputs sejam dados, novos serão descobertos e o perfil traçado pela máquina será mais seguro; * No período de \_\_\_\_, novas entradas devem ser rastreadas e inseridas no programa. | **Offline Evaluation**  Methods and metrics to evaluate the system before deployment.   * Checagem constante de novos dados relacionados à doença; * Observação de bases de dados pública do sistema de saúde local; * Acompanhamento direto de notícias e realização de pesquisas diversas para comparação dos índices anotados. |  | **Features**  Input representations extracted from raw data sources.   * gender: "Male", "Female" or "Other"; * age: age of the patient; * hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension; * heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease; * ever\_married: "No" or "Yes"; * work\_type: "children", "Govt\_jov", "Never\_worked", "Private" or "Self-employed"; * Residence\_type: "Rural" or "Urban"; * avg\_glucose\_level: average glucose level in blood; * bmi: body mass índex; * smoking\_status: "formerly smoked", "never smoked", "smokes" or "Unknown"; * stroke: 1 if the patient had a stroke or 0 if not. | **Building Models**  When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?   * A medida que novos inputs sejam dados, novos serão descobertos e o perfil traçado pela máquina será mais seguro; * No período de 6 meses, novas entradas devem ser rastreadas e inseridas no programa. |
|  | **Live Evaluation and Monitoring**  Methods and metrics to evaluate the system after deployment, and to quantify value creation. | * Os dados serão constantemente monitorados por gráficos e métodos de busca (valores NA, outliers, etc); * As bases de dados públicas que primordialmente são utilizadas, serão do mesmo modo monitoradas para eventuais atualizações do modelo. |  |  |