**Capstone Project**

**Project proposal**

**1.** **Group description**

**1.1.** Group name

|  |
| --- |
|  |

**1.2.** Students names, background and target industry if any

|  |
| --- |
| Brandon Ryu  Douglas Hilton  Stephen Kita |

**1.3.** Group structure: roles and responsibilities

|  |
| --- |
| EDA – split into thirds: admissions, lab results, surgical  unsupervised learning – split into k-means, PCA, and LDA  first model – split into random forest, logistic, and boosting  validation and refinement – present to each other and continue modeling  boom. |

**2. Why** do we want to develop a data science project?

**2.1 Objective**: what problem do you want to solve? What questions are you trying to answer? How will you **measure the success** of your analysis from a business/user perspective?

|  |
| --- |
| We are trying to predict congestive heart failure with a machine learning model. Alternates are stroke, flu, or kidney failure  measures of success:   * accuracy, precision, sensitivity, ROC curve * compare against clinical diagnosis criteria (vs hypothetical ROC curve backtracked from clinical data) * earlier diagnosis and something-something disease burden |

**2.2.** **Scope** of application: what population and timeframe will your analysis/model be applied to or used for?

|  |
| --- |
| description of the database |

**3. How** do you translate the objective and scope in terms of data?

**3.1.** What **dataset**(s) do you plan to use? Initial description: source, granularity, number of observations, variables list…

|  |
| --- |
| Description of the data tables |

**3.2.** What **data treatment and analysis** do you plan? Data Aggregation, target variable definition, tools, analysis/machine learning, etc.

|  |
| --- |
| index on patient ID, aggregate vital measurements, lab results, transfer to surgery and outcomes  target congestive heart failure. start with binary heart or not, move to binary heart failure or not, end with multiclass.  split data on demographics, train with different data and compare feature significance and performance  general supervised classifiers, support vector machine, random forest, gradient boosting  general unsupervised clustering (PCA and K-means)  K-Means, PCA and LDA to reduce feature space |

**4. Project plan**

Please submit a project plan proposal broken down by a few significant steps. Plan at least three meetings with your stakeholders.

- **Kick-off meeting**: schedule a 30 minute meeting before project declaration in order to approve the project proposal.

- **Milestone 1**: schedule 30 minutes with your stakeholders to present initial results and insights and to validate any assumptions or definitions needed to move forward.

- **Milestone 2**: schedule 30 minutes to go over the final results and proposed presentation before the final presentation in front of the whole team.

EDA – 1 week

Early models – 3 days

Final models – 5 days

If you think additional discussions will be required, feel free to add secondary milestone(s) in your project plan.

**Project plan and schedule examples**

(create and use your own template)



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **September** | | | | | | | | | | | | | | | | | |
|  | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| **Kick off** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Project declaration |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| EDA |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Unsupervised Learning |  |  |  |  |  | push > |  |  |  |  |  |  |  |  |  |  |  |
| **First Model** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Model Validation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Model Refinement |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Final Model** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Delivery** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |