

Team name: BEAMING

Team members: Dilay Ozkan, Nghia Nguyen, Benjamin Torres, Malavika Mampally, Egemen Elver

Date: 04/12/2024

Team roles for this report (write down name):

Facilitator(s): Benjamin Torres

Recorder(s): Dilay Ozkan, Nghia Nguyen

Deliverer(s): Malavika Mampally, Nghia Nguyen

Planner(s): Malavika Mampally, Egemen Elver

Team Contact: Dilay Ozkan (dilay@unc.edu)

- **Describe briefly what the main goal of your team is (so the peer reviewer has some context). E.g. we are working on image classification for blah de blah. Our goal is blah de blah etc. In the initial part of the semester before your proposal it is ok to put down “we are still coming up with ideas on team project”.**

Our team's goal is to check how well each model does based on accuracy, precision and other classification metrics. We made note of what variables were useful, what were not. We wish to keep the data consistent in terms of data cleaning and organization.

I. What was done during the report period regarding the project:

We each picked a different modeling technique from last week and discussed our findings. Malavika noted cross-validation results of KNN which were not as good as other results. While running a QDA model, multicollinearity was encountered and we decided on correlation threshold that should be applied to eliminate highly correlated

variables. Egemen worked on LDA, that showed some improved results compared to KNN and then Benjamin walked us through his method of applying XGBoost and Tabnet to get the highest accuracy so far. (96%). Dilay has been working on Naive Bayes classification method to see if does any good in differentiating between a phishing and legitimate data. Finally, MLP results were discussed by Nghia which seemed to work well with our dataset.

II. What were obstacles faced if any in working on the project?

- The dataset exhibited multicollinearity due to which we had to create conditions based on correlation and variance influence factor to eliminate unwanted variables.
- Last week was busy for all members due to the Midterm. We could not set aside time to meet up and discuss our workings on the project.
- Since it was the week of 2nd midterms, other courses also had their exams and projects due which served as an obstacle in the project's progress.

III. What is the plan for the next reporting period including what each team member is planning to work on. Describe goals and potential timelines.

1. To maintain uniformity, we decided to use the same training and testing dataset to fit all our models. This is to avoid bias (not the statistical one) due to different seed values.
2. Write down a detailed report on each modeling technique used to classify the data, organize the outcomes to facilitate comparison and selection of the best
3. We will begin adding content to our presentation slides based on our final model.