* 3-5 minute video telling a short story of your work
* Can use real or animated characters
* Should have nice illustrative figures and compelling story
* NOT reading the script in front of a camera!
* Each team submit an URL to your video on Collab due on **11/15**
* Write a script and a storyboard to help you plan the story
* Film your story (with smartphone camera) with multiple takes
* Edit your video using any media editing software and tools
* Give credits to all of your sources (information, 3rd party media) and include the following in the credit: *This work has been a part of the “Machine Learning for Virginia” project at the University of Virginia in Fall 2019.*
* Upload to YouTube under a public or only access by URL
* **Content**: A clear statement of purpose or theme and is creative, compelling and clearly conveyed. Show strong understanding and in-depth analysis about the topic
* **Organization**: Media shows high degree of attention to logic and reasoning of points. Unity clearly leads the audience to conclusion and stirs thought regarding the topic.
* **Discussion** and **Analysis**: Clearly discuss what you have learned from doing the project. Main points are well developed with high quality and quantity support.
* **WOW factors:** Amazing story with lots of creativity. The approach is innovative and visually appealing.

Introduction:

* introduce what dataset we used
  + This is a Virginia Beach Satisfaction survey, it contains 3028 rows and 69 columns. This dataset was published on March 3rd, 2018, while the survey itself was conducted in 2015. The purpose of the dataset is to provide timely and accurate city information to increase government transparency. The majority of the questions have categorical answers: such as satisfaction, ranging from very dissatisfied to very satisfied; agreement, ranging from strongly disagree to strongly agree; and yes-no questions.
* What we’re trying to predict
* Show graphs/images of the information from the dataset
* Show demographic information?

Slide 1:

<slow down>

This is Bob. He is a 40 year-old male living in Virginia Beach for the past 3 years, and he is unhappy with his life.

Slide 2:

Maybe his dissatisfaction comes from the fact that he is a software engineer who works over 70 hours per week, or because of his newborn or he is simply having a mid-life crisis.

Slide 3:

Instead of dealing with his problem directly, he decided to see if people who are in similar situations are also unhappy.

While ---Bob stumbled across some survey from Virginia Beach, which included around 50 questions on general life satisfaction, and demographic data of the responder.

Slide 4

Data set

There were 50 questions in the survey and the survey responses were all categorical variables with 5 categories ranging from very dissatisfied to very satisfied. Bob predicts the overall satisfaction score- which was calculated by assigning numerical values to the responses in the survey and modeling them using Machine Learning techniques.

Slide 5

Methods: Talk about the rmse

He started with good-old linear regression.

* Linear regression, rmse: 0.1315, which is not good given that the standard deviation is 0.169
* Polynomial Regression: rmse 0.16, overfitting, regularized it was still bad
* SGD Regression: rmse: 0.129 Then, he used polynomial regression and found it was similar to his life

The

He had a glimpse of hope, when he learned about random forest in his machine learning class (put Nguyen’s face). However, the fate was cruel. It was also shit.

Slide 6

Scaling the response variable, hyperparameter tuning, Polynomial Spline and regularization .Even the fanciest techniques with extensive preprocessing were no match for this dataset.

Slide 8

Through this analysis, Bob learned that it is important to have data with more quantitative features. A lack of quantitative features can cause the algorithm to perform poorly especially for regression, thus limiting the capabilities of the model.

Bob also realized that it is crucial to have data features with high correlations to predict the label with the lowest possible error. With the Virginia Beach dataset, the correlations weren’t above 10%, which limited the capabilities of the model to perform well. That’s what happened on a granular level.

**Further Steps**

1. Classification of Very Dissatisfied to Satisfied

Instead of predicting the satisfaction score, Bob also wants to try classify where people are very dissatisfied to very satisfied and curious to see if the performance i s good the classification model

1. CNN

Maybe, he wants to pursue digging deeper into the regression and try neural networks.

**Slide 9**

**Conclusion**

Stepping back, it came to his mind that human emotion could not be predicted by a simple “dataset” or a “survey”. There wasn’t any one feature that could predict life satisfaction,

There are many many factors that can contribute to life satisfaction that were not captured by the dataset. Whether he was able to predict the life satisfaction or not, Bob learned a important life lesson.