Dating App Spam Filtering Models

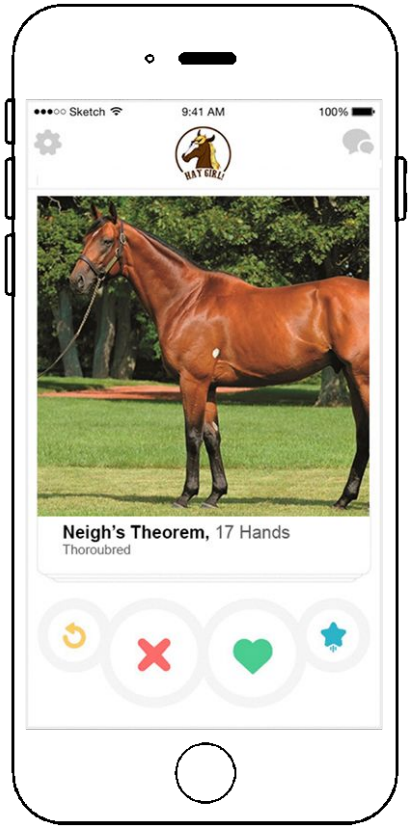
Meet Virginia

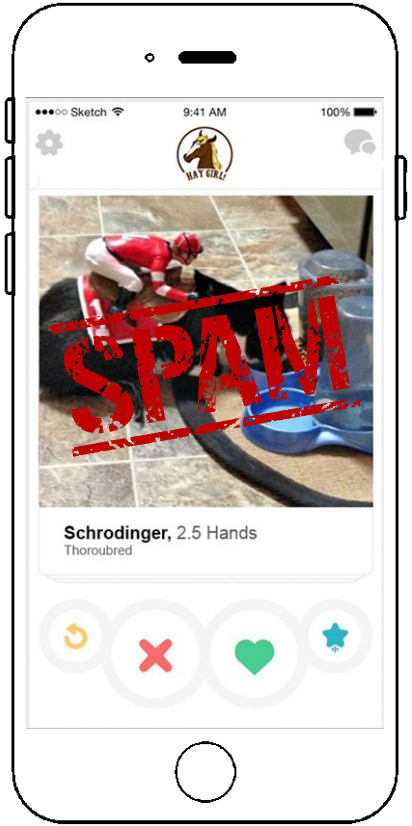
(and her horse: Neigh's Theorem)

- Former Competitive Barrel Racer
- Amature Horse Breeder
- Looking for suitable breeding stallions
- Developed Dating App to Pair Breeding Horses:

HAY GIRL!

Filtering out Spam Profiles from Horse Dating App: “Hay Girl!’





Consultants

Disha Mac

Samuel Pepper

Dan Chernoff

Ryan Maciej

NEARESTNEIGHBORS

DATA CONSULTANTS

Agenda

1. Exploratory Data Analysis Findings
2. Preprocessing & Transformations
3. Initial Models
4. Optimized Model
5. Comparing Models
6. Summary

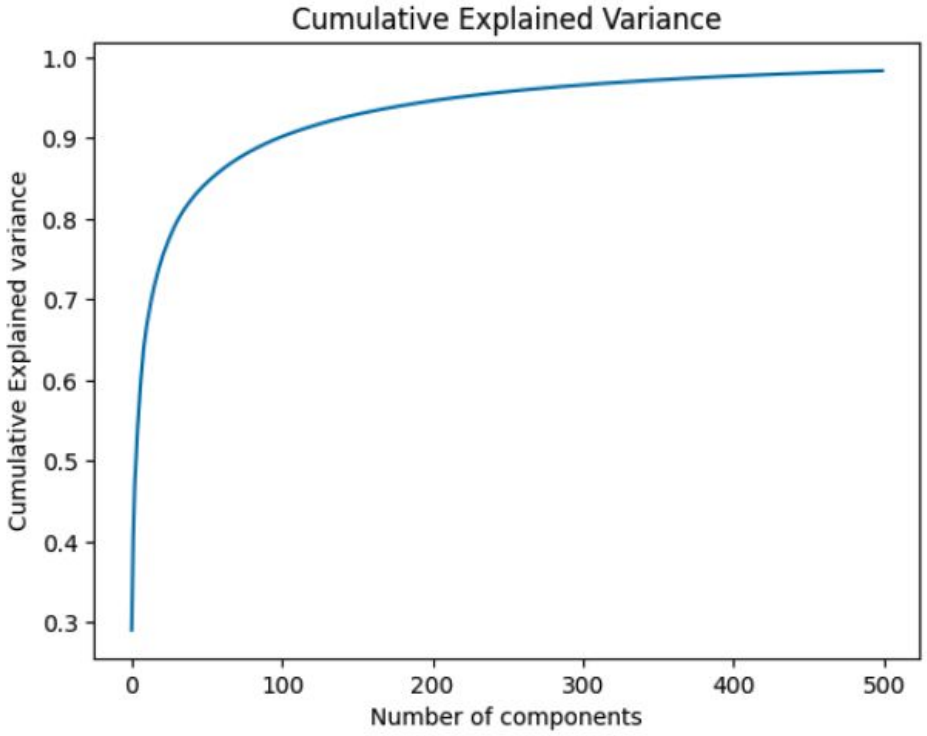
According to a **Gallop** Poll:

100%

of horses love to run.

Can Vaguely distinguish a horse from Mean Images by class

Aprox. ¼ of Features (500) make up 98% of Variability



The Mane Event:
Building a Stable Model

Not this type of Stable Model

Data Transformations For Initial Models

- Binary Representation of the Response
- Flatten both Training and Test
- Scale the Training and Test
- Create Features Using PCA

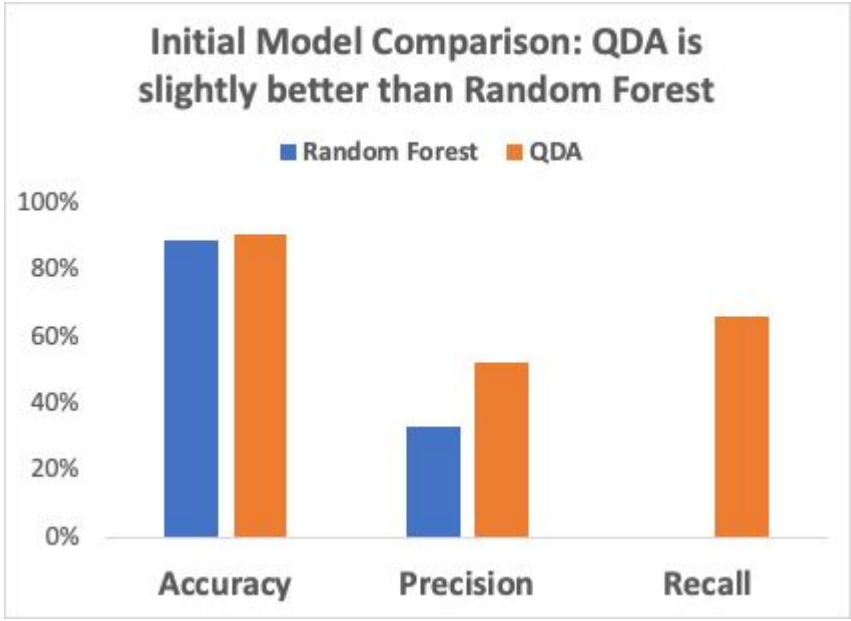
Initial Models Used to Establish Baseline

Model 1: Random Forest for classification

- These are multiple decision trees that predict if the image is of a horse by learning simple decision rules inferred from our data set

Model 2: Quadratic Discriminant Analysis for classification

- We model the distribution of our predictor data for both horse and non-horses

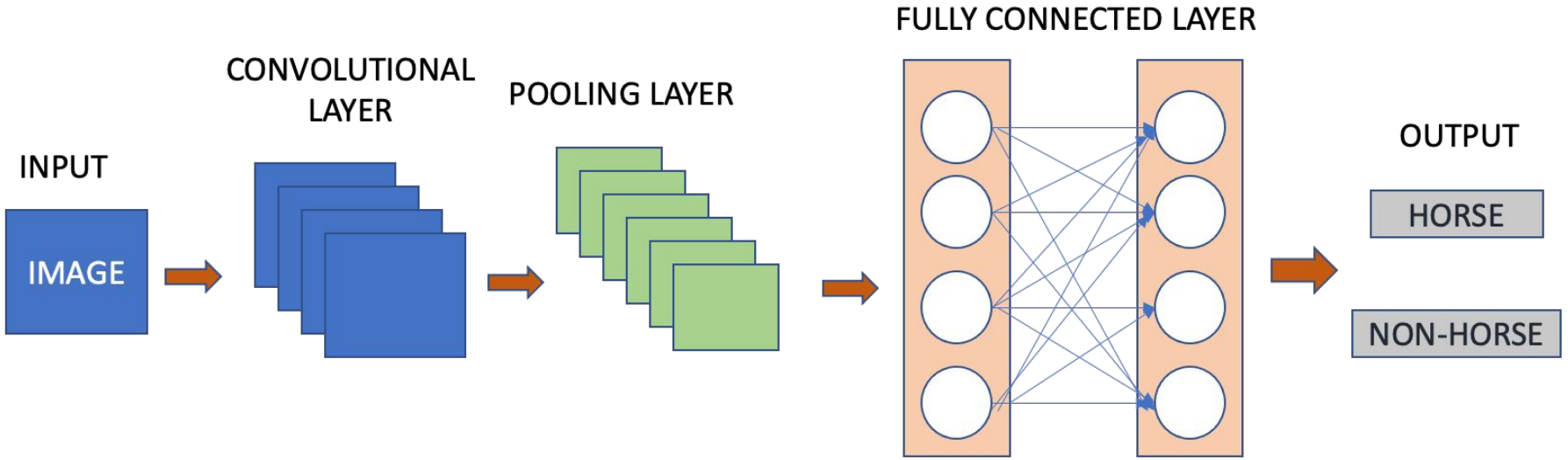


Data Transformations For Optimized Model

- Binary Representation of the Response
- Categorical Representation of Response
- Normalize Training and Response Data

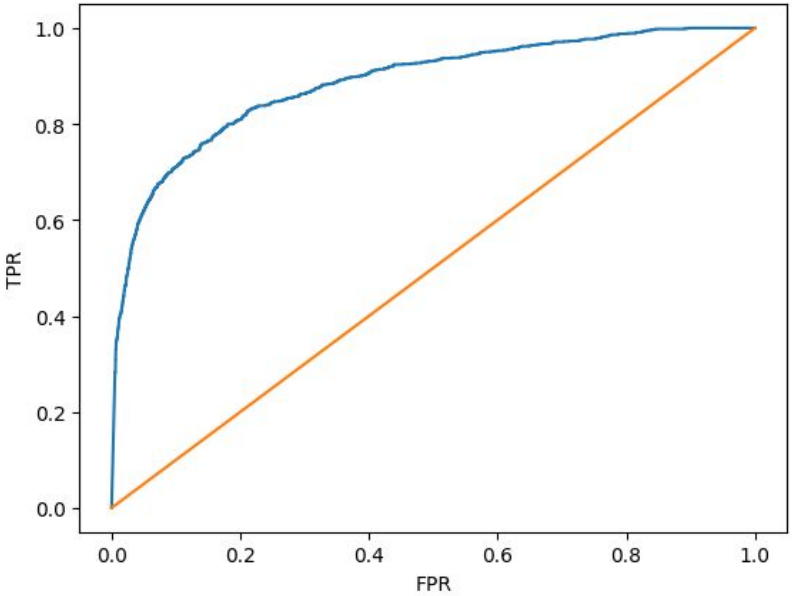
Optimized Model for Better Predictions

Sequential Convolutional Neural Network to get better classification

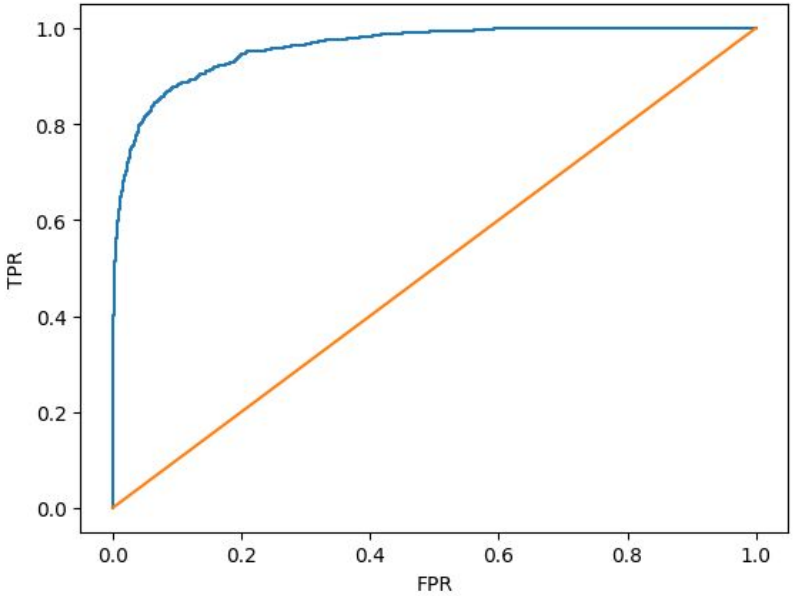


ROC for Optimized Model is Better than Initial

ROC Initial Model (QDA)



ROC Optimized Model



Optimized Model Better Overall

Metrics Initial Model (QDA)

Overall Accuracy	90%
Precision	52%

Metrics Optimized Model

Overall Accuracy	95%
Precision	85%

An eQuestrians?