PREDICTING JOB APPLICANT ATTRIBUTES FROM INTERVIEW DATA

intro to Data Science final project

The Problem

During an interview, we can observe various traits such as attentiveness, calmness, confidence, communication skills, and creativity. **However, predicting a candidate's persistence, a crucial trait for many roles, is challenging.** Can we leverage machine learning to predict a candidate's persistence based on their responses and observed traits during an interview?

S.M.A.R.T Goal

The Goal

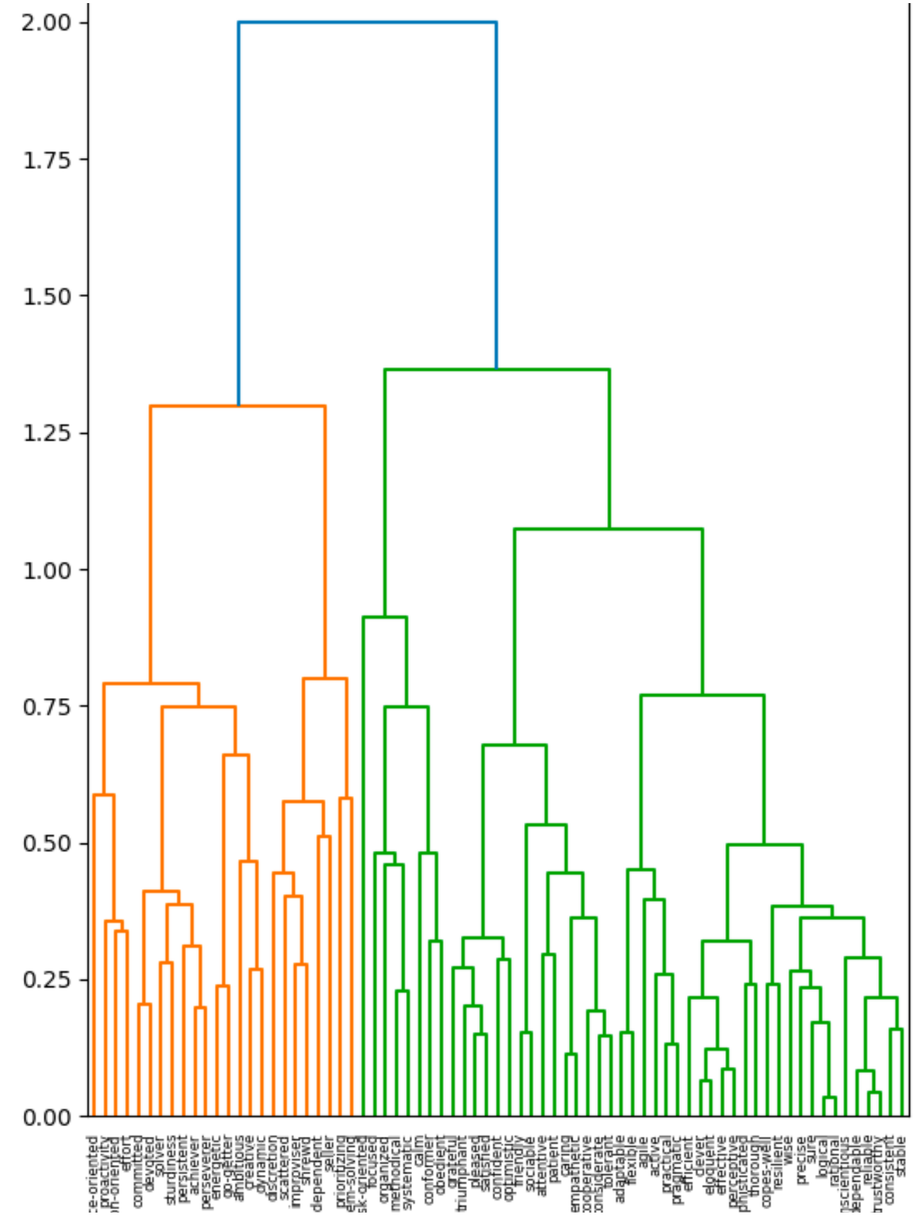
To develop by the end of the course a machine learning model that can predict at over 90% accuracy complex personality traits such as persistence, based on first impressions from an interview

How do we get there?

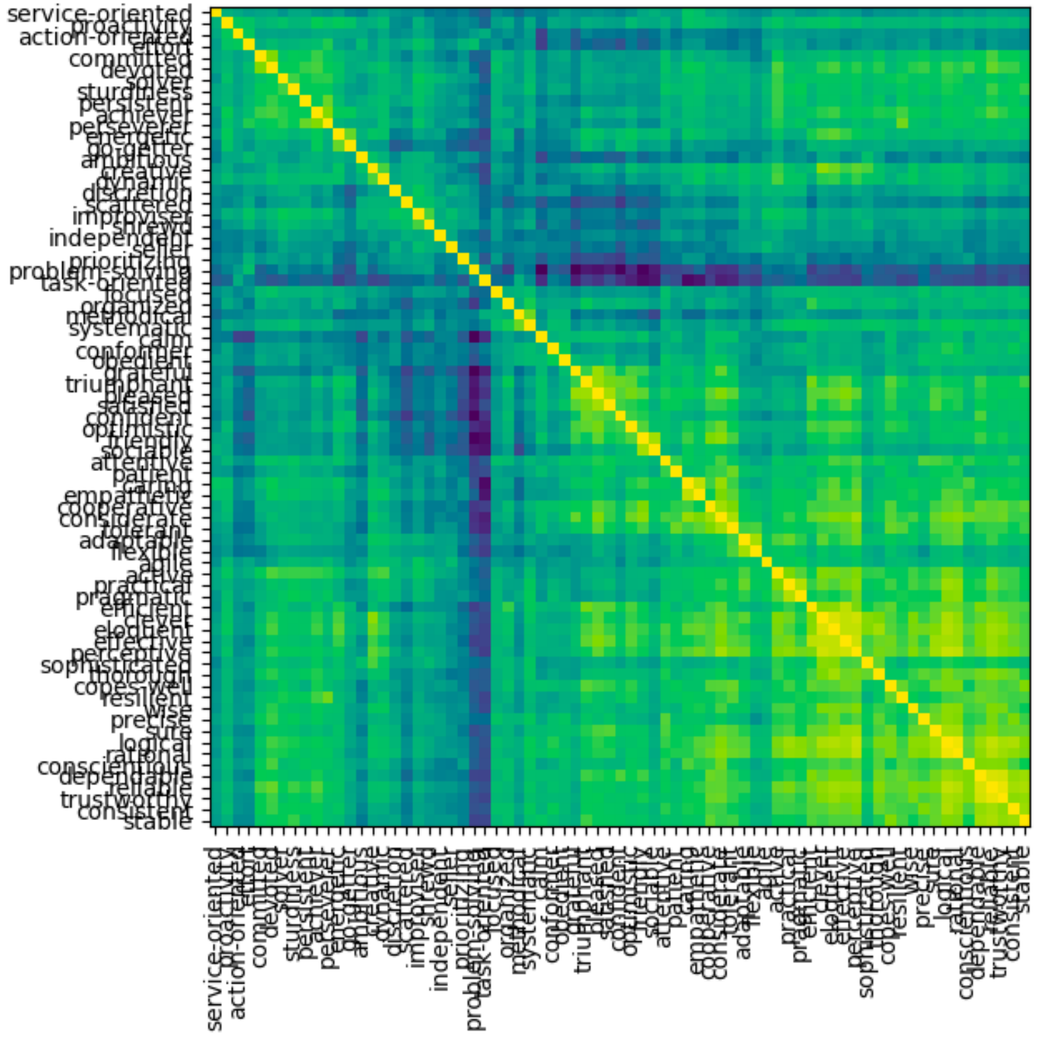
By using advanced machine learning methods like XGBoost and Polynomial Regression, we can train a model on interview data to accurately predict complex personality traits

EDA

Hierarchical clustering using Ward's linkage



Correlated features clusters



Data preparation

Manual differentiation	Correlation matrix	Variance Thresholding	Multi-collinear features clustering	Split data
We manually split the dataset between simple and complex traits 8 complex traits	Show basic correlation between traits	Prune features that are relatively invariant Dropped 9 highly invariant features	Unify groups of highly correlated features Left with 40 independent features	We used a 80-20 split The split kept all instances of specific person in the same group (test/train)

HOW WE DO IT: MODELING

Gradient Boosted Decision Trees (XGBoost)

A collection of decision trees that learn from past mistakes, making it more accurate with each iteration

Polynomial regression

Models the relationship between the independent and dependent variables as an equation capturing nuanced relationships between variables.

EVALUATION XGBOOST

Traits	Accuracy	MSE	Test/Train Accuracy Ratio
Persistent	94%	0.0027	94%
Considerate	100%	0.0004	100%
Seller	96%	0.0021	96%
Committed	93%	0.0043	93%

EVALUATION POLYNOMIAL REGRESSION

Traits	Accuracy	MSE	Test/Train Accuracy Ratio
proactivity	96%	0.0025	96%
independent	99%	0.0017	99%
go-getter	100%	0.0005	100%
organized	93%	0.0028	97.8%

RESULTS

HIGH PREDICTION RATE

POLY > XGBOOST

FURTHER INVESTIGATION

How would our model perform using "real-world" data?

Use publicly available HR datasets

or build one ourselves

Do you have any questions?