Krannert IBM Watson Personality Insights Prototype

Alex Chirayath, Aakriti Garg, Russ Gentile, Abhisek Gupta, Samuel Hiatt, Devesh Kumar Singh

Purdue University Krannert School of Management

achiraya@purdue.edu; garg76@purdue.edu; rgentil@purdue.edu; gupta362@purdue.edu; hiatts@purdue.edu; singh349@purdue.edu

Abstract

The project assesses the use of IBM Watson to identify an equitable mix of incoming graduate students. In this first phase, quantitative estimators of a student's future performance in Krannert are derived from the application submittal. Estimators are an additional input to Purdue's holistic admission review process which considers many different elements. This research holds importance by demonstrating an innovative data-informed decision support system for the admissions process, while considering unintended consequences of a predictive modelling approach.

Introduction

The IBM Watson™ Personality Insights service provides an Application Programming Interface (API) that enables applications to derive insights from social media, enterprise data, or other digital communications. The service uses linguistic analytics to infer individuals' intrinsic personality characteristics, including Big Five, Needs, and Values, from digital communications such as email, text messages, tweets, and forum posts. This prototype uses Watson to derive personality insights from a student application to Krannert School of Management.

Krannert, with IBM.

"… the risk of investing too late in smart machines is likely greater than the risk of investing too soon."
— Gartner Group, 2014

Why Watson at Krannert?
- Krannert research interests in cognitive augmentation used for improved outcomes, in business processes and personal experiences, aligns with IBM Watson technology
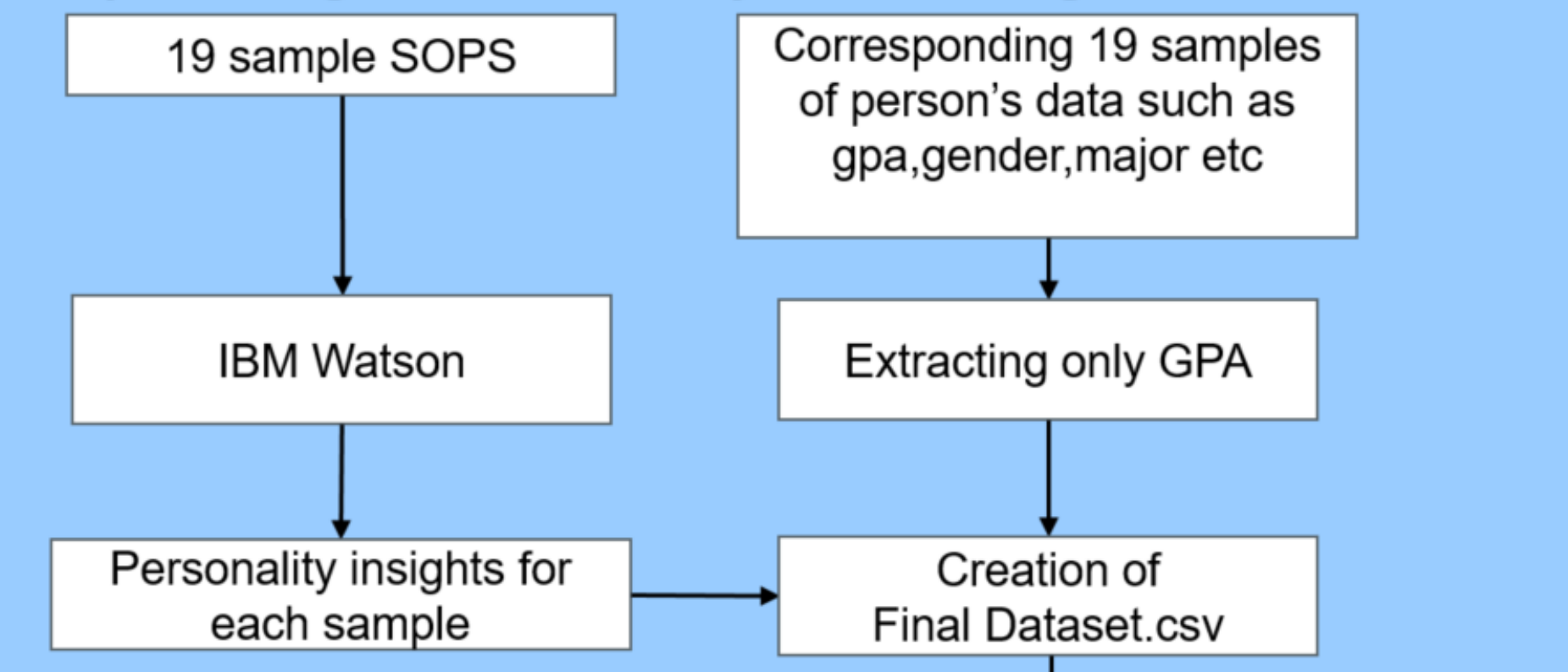
How we are using Watson:
- Add personality insights input to admissions process
- Predict student success
- Promote cognitive diversity

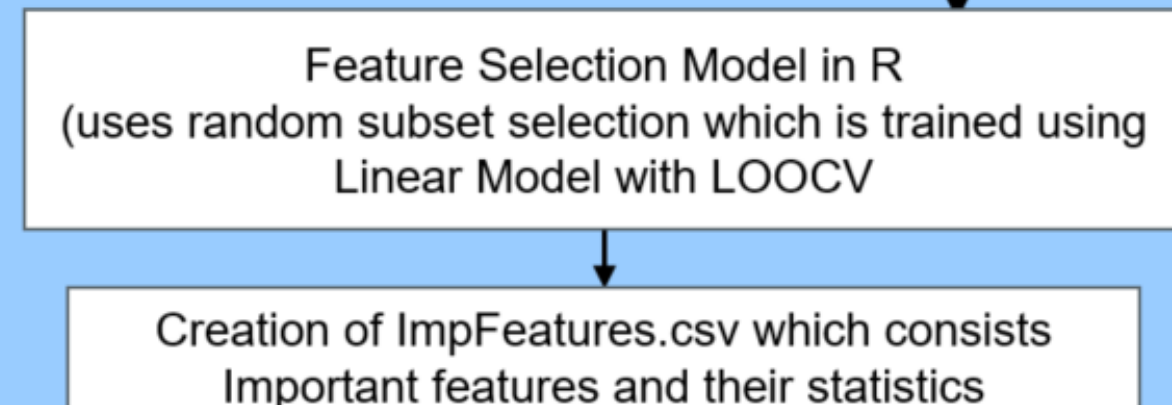**IBM WATSON COGNITIVE TECHNOLOGY IS PROVEN**

Methodology

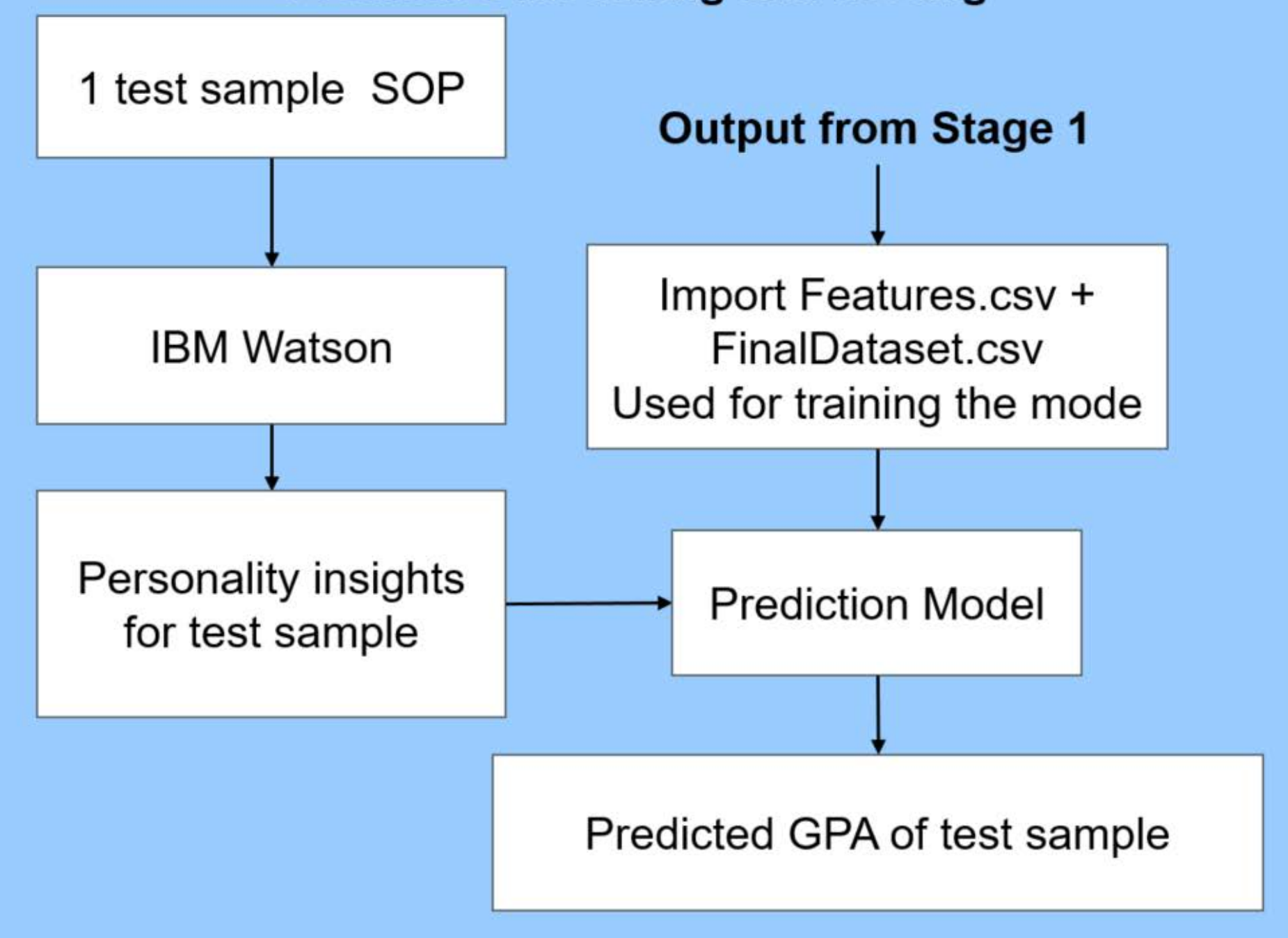**Preprocessing of data done in Python and using IBM Watson APIs**

Stage 1 – Data Preparation

- 19 sample SOPS
- Corresponding 19 samples of person's data such as gpa, gender, major etc
- IBM Watson
- Extracting only GPA
- Personality insights for each sample
- Creation of Final Dataset.csv

Feature selection done in R
- Feature Selection Model in R (uses random subset selection which is implemented using Linear Model with LOOCV)
- Creation of ImpFeatures.csv which consists Important features and their statistics

Input to Stage 2

Predictive Modelling and Testing — Stage 2 – Predictive Modelling
- 1 test sample SOP
- **Output from Stage 1**
- IBM Watson
- Import Features.csv + FinalDataset.csv Used for training the mode
- Personality insights for test sample
- Prediction Model
- Predicted GPA of test sample

RATIONALE FOR CHOOSING A PREDICTIVE MODEL

Upload SOP of potential students

IBM Watson Analysis + Feature Selection + Prediction Model = Predicted GPA

Because of the proven capability of IBM Watson Personality Insights, the results are expected to be a good predictor of GPA. Future iterations of the application can support other metrics including predicted post-graduate outcomes.

FULL AUTOMATION USE CASE – IBM WATSON AS MAIN TALENT IDENTIFIER
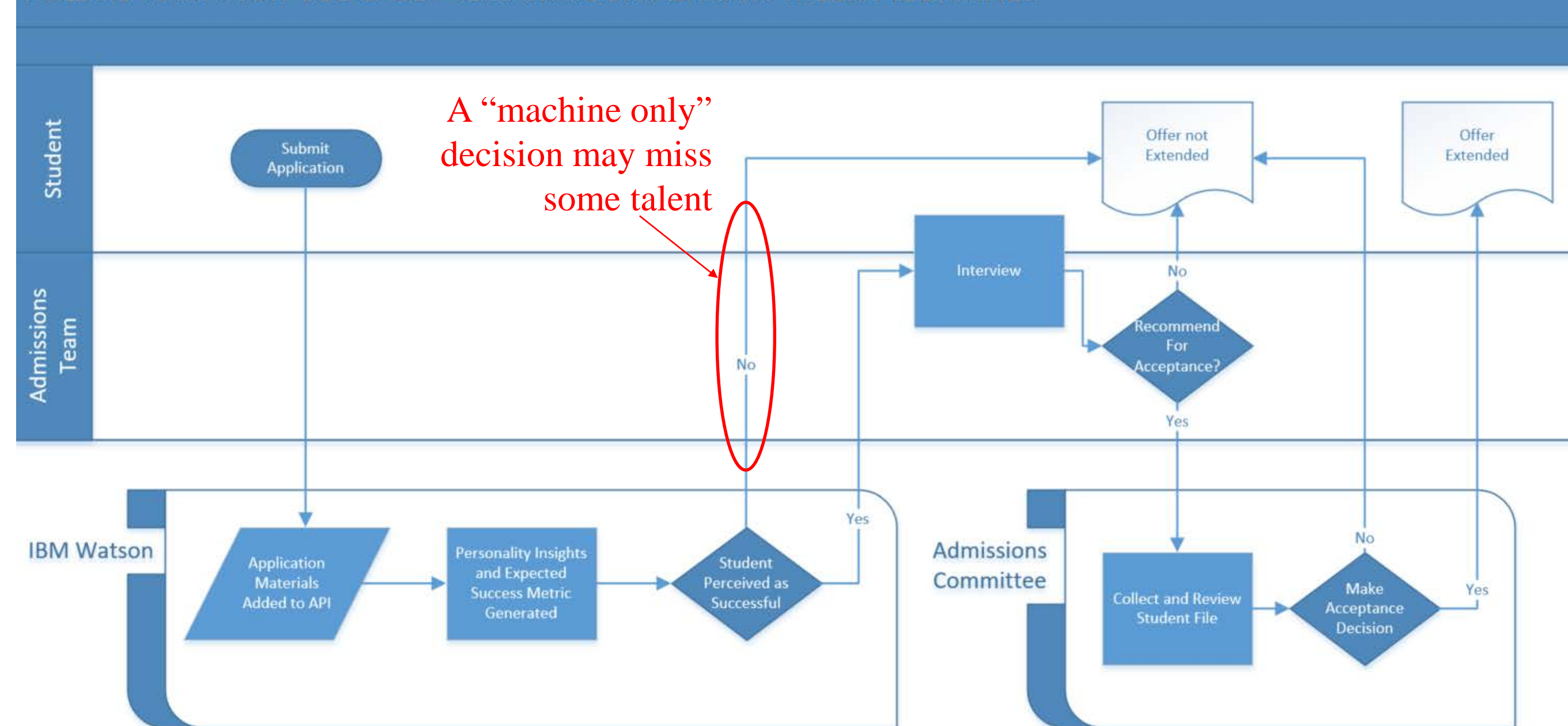
A "machine only" decision may miss some talent

### STAFF AUGMENTATION USE CASE – DECISION SUPPORT METRICS
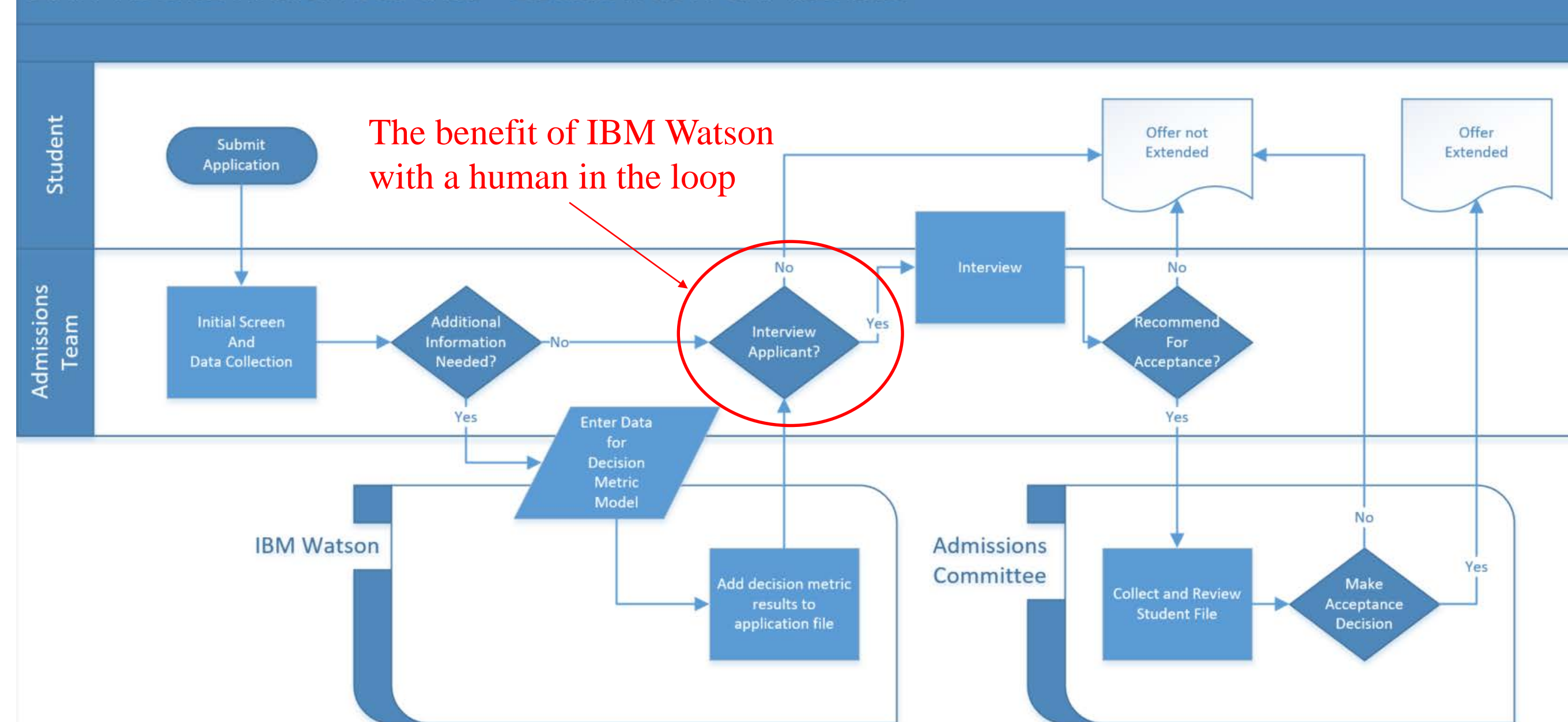
The benefit of IBM Watson with a human in the loop

THE TEAM ADDRESSED SEVERAL ISSUES DURING DEVELOPMENT

Issues	Solution
IBM Watson accepts junk words for analysis	System uses a dictionary to filter the text
IBM Watson can't read PDF files	Python code converts PDF files into a text file
Lack of data for model creation	Ran a higher number of simulations (10K) for accuracy
Correlation in some of Watson's 47 features will impact the accuracy of the model	Used random subset selection method which revealed 14 features with higher significance in the predictive model
Dataset too small to run complex models such as Artificial Neural Networks, Higher order regression, Decision Trees	Implemented Linear Regression and Lasso CV model
Dataset too small for efficient cross validation (K-fold validation / to implement Bootstrapping)	Used LOOCV method to cross validate answer

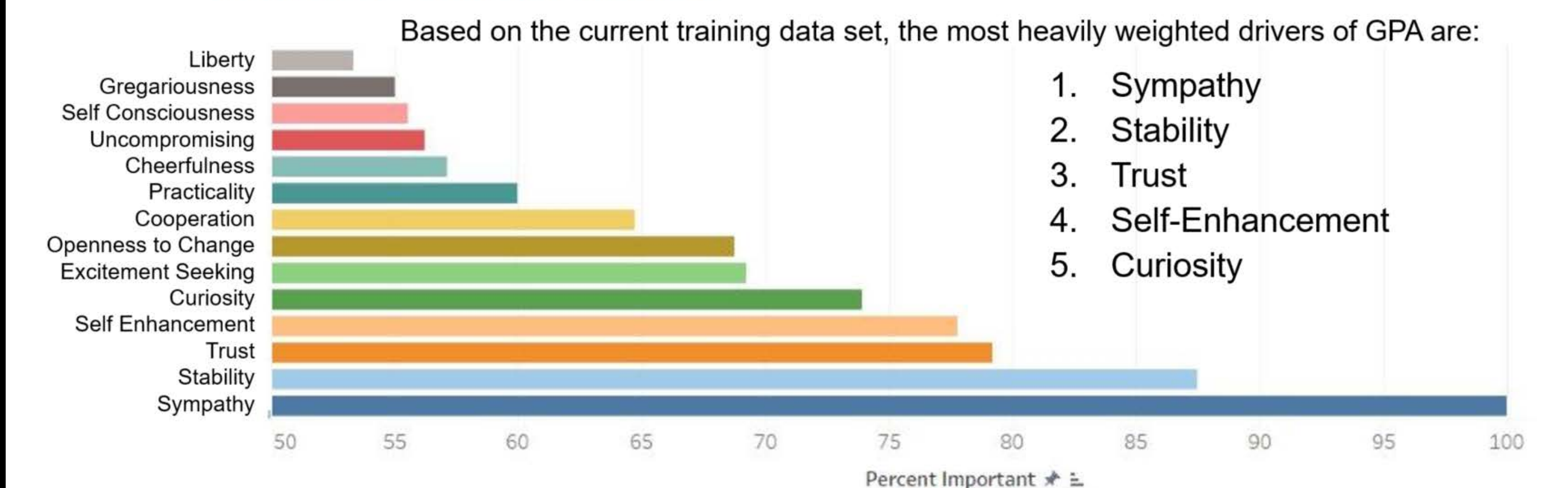Prototype of application screen viewed by an admissions team member:

Please Upload Your files
SELECT FILE(S)

Predicted GPA result:
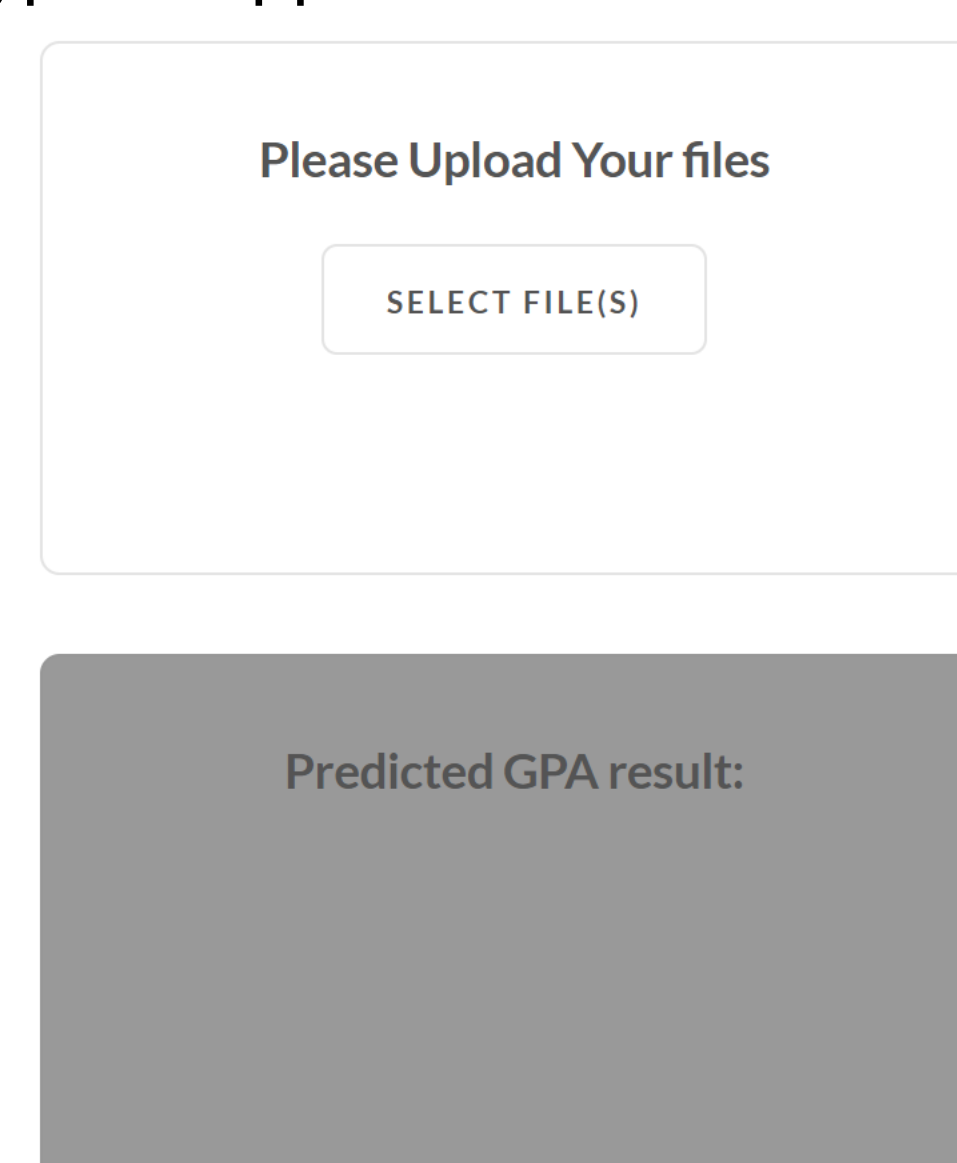
Score on top 5 Traits
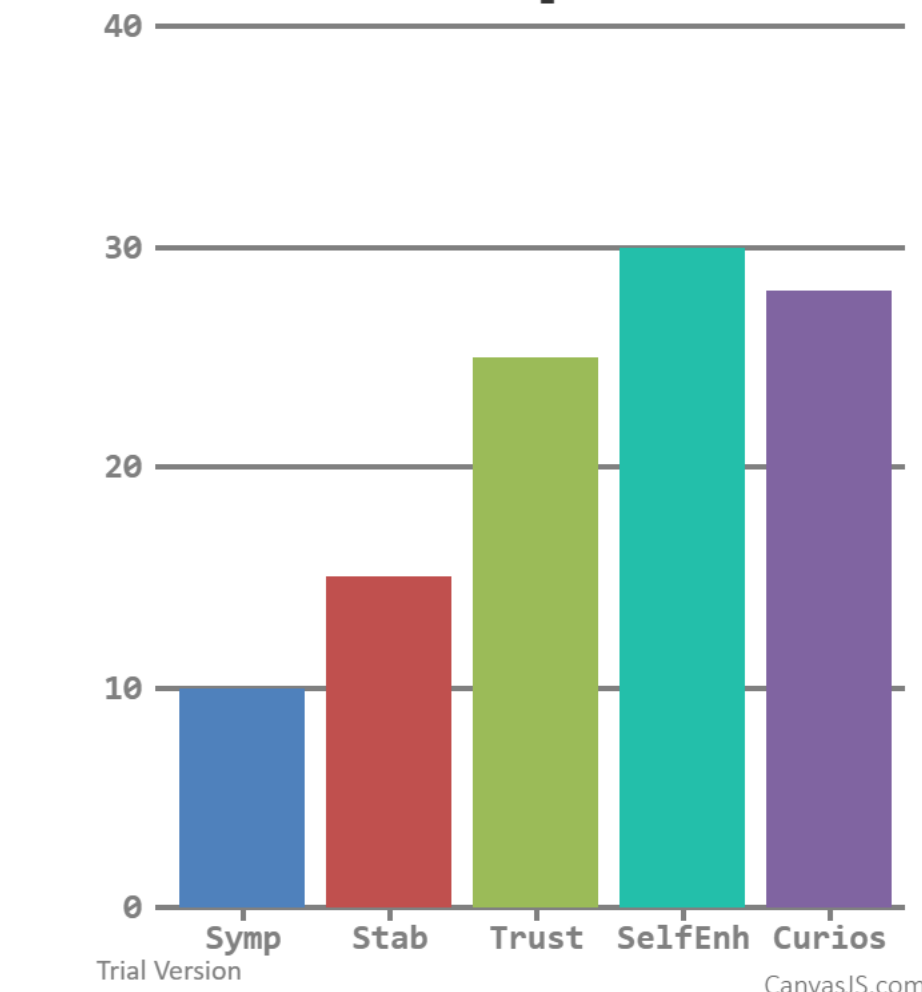
Results

Of the 47 personality insights offered by IBM Watson, 14 were determined to be the most significant in the prediction of future student GPAs.

PERSONALITY INSIGHTS FROM PILOT PROJECT – SPRING 2017

Based on the current training data set, the most heavily weighted drivers of GPA are:

1. Sympathy
2. Stability
3. Trust
4. Self-Enhancement
5. Curiosity

Students that posses a combination of these important features are examples of candidates that would be expected to grow into a successful 21st century business person during their time at Krannert School of Management.

Decision Model

Feature Selection
- Choosing random sets of features due to the data constraints seemed intuitive for several thousand simulations
- The models are cross validated using LOOCV model for each simulation
- We chose instances and features from these with low p-values(<0.1) to find the more significant features

Model Selection
- Small data size led to use of simple models such as Linear Regression, Lasso Regression and Ridge Regression
- Lasso Model was selected as the final model since
  1. Linear Regression could lead to elevated standard error in the parameters
  2. Ridge regression cannot zero out coefficients; thus, you either end up including all the coefficients in the model, or none of them.
- Lasso will also tend to have smaller variance giving an overall more accurate prediction than the standard least squares estimate, particularly in the presence of small samples (Tibshirani, 1996)
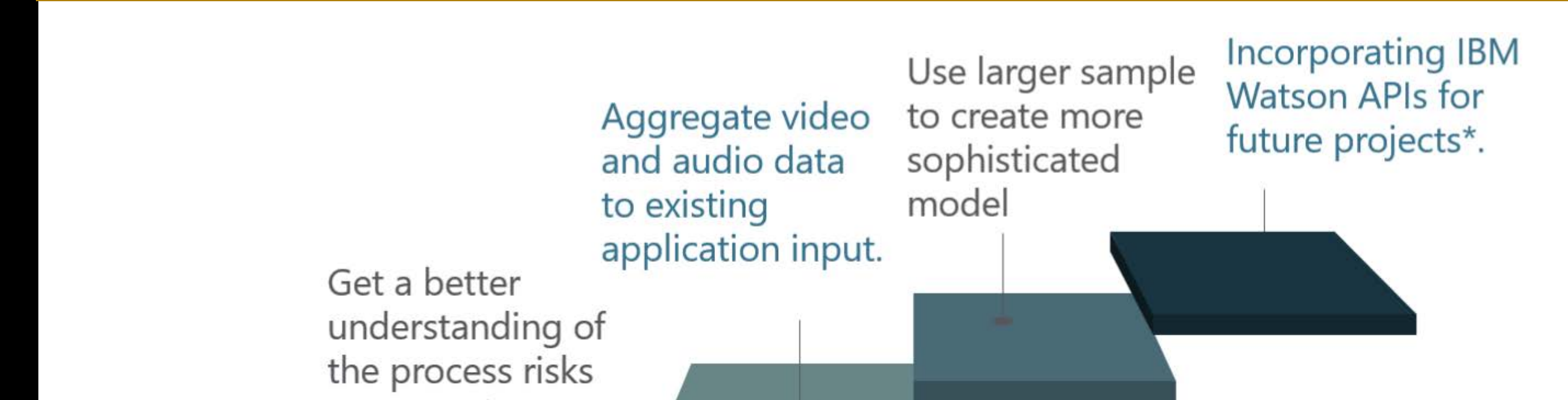
Conclusions

IBM Viability
- Personality insights can be used to generate predicted GPA scores for applicants
- Scores can be used by admissions team as a decision support system in evaluating candidates
- Analytic solution is determined through predictive modeling

Benefits to Krannert Admissions
- Aid to the process
- Predicting GPA and analyzing student personality traits
- IBM Watson can be used to add more data points to the decision making
- In the future, model can be expanded to provide additional inputs such as predictions for post-graduate success metrics

Future Scope

Get a better understanding of the process risks

Aggregate video and audio data to existing application input.

Use larger sample to create more sophisticated model

Incorporating IBM Watson APIs for future projects*.

* Explore IBM Watson to create a Purdue Chatbot.

* Explore IBM Watson Analytics for Krannert's Student Managed Investment Fund.

Acknowledgements