Data Edvantage

Data Analysis of Majulah Junior College's Internal Physics Exams

Class: IBM IMVAI-1502

Team: 4

Team Members: Rahbeiatul Fazeria, Melodie Tay, Quek Wee Tong, Loretta Yong, Tan Siang Khim

Agenda

- Introduction
- Planning Analytics
- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics
- Summary & Reflections

Introduction

Project Objectives

To help Physics educators in Majulah Junior College:

Be more effective in setting better exam questions

Improve their overall teaching methods and learning journey for their students

Tools and Applications

Collaboration and project management

Descriptive statistics

Data compilation, refinery

AutoAI for building and deploying ML models

Planning Analytics

Our Target Users

"Who hired us?"

Majulah Junior College

Physics Department

"What are we hired to do?"

Analyse past exam data

Recommend actions to take

"Who do we work with?"

Level heads, who coordinate exams

Teachers, who help set and vet test items

Problem Statement

"How might we help Physics educators in Majulah Junior College be *more effective* in setting *better exam questions*?"

Motivations

Set better exam papers

Improve teaching & learning for their students

Goals

Gain insights from from past exam data

Use these insights improve processes in assessment and teaching

As-Is Scenario

How an exam paper is set?

Design paper	Setting	Peer vetting	Team vetting	HOD approval
How many questions? What topic to test? How many questions of each topic, etc…	Teachers received guideline, sets the question	Teachers assigned a peer vetter 1st round of vetting	Group of teachers sit down together to look through entire paper	Head of department final vetting

As-Is Scenario

How are exam questions set?

Teachers' thinking process

- Design from scratch
- Modify from past questions vs pure inspiration

Labelling test items

- Topic & learning outcomes (according to A-levels syllabus)
- Nature of question -
  - Calculation, graphical, conceptual, or mix?
- Predicted difficulty level: Easy, Moderate, Difficult

Collating item features

- All of these data are collated in a table called Table of Specifications (TOS)
- Used for reviewing test as a whole

- Exams statistics are generated for every MCQ paper
- Test statistics: Mean, quartile scores, test reliability score*
- Item statistics: Stats for choices, % correct, item reliability score*

- Data collected for every exam
- Post-mortem done only on single exam
- But no process/bandwidth to evaluate them over long term
- Data also not used to influence future setting decisions

As-is Scenario

What happens after exam?

Difficult to improve effectiveness in question setting

No reliable way of predicting difficulty and reliability of whole test

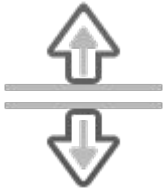
Pain Points

No way to measure and verify the 'hunches' that teachers have towards students' weaknesses

Unable to distinguish topics or question types that tend to be less reliable as test questions

Let's uncover new insights!

Purpose of Exams



Assess learning, discriminate students ability

Provide feedback to both students and teachers

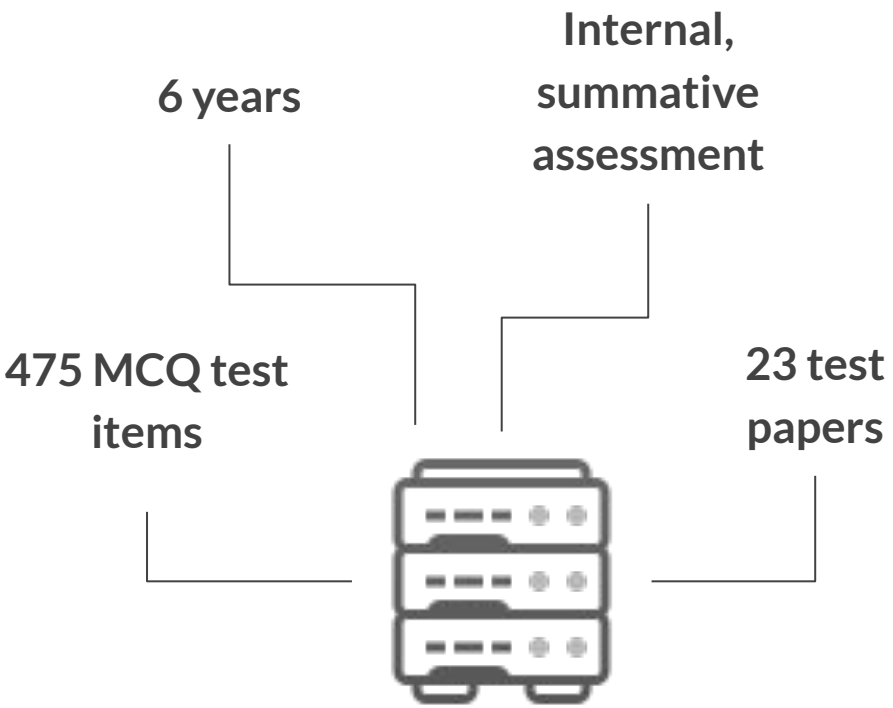
Guidance for subsequent teaching / learning strategies / intervention

Qualities of a good MCQ Exam Paper

- Objective: To discriminate students' level of competency
  - A test paper comprises exam questions of varying context and nature
  - Test, as a whole, must be reliable and accurate

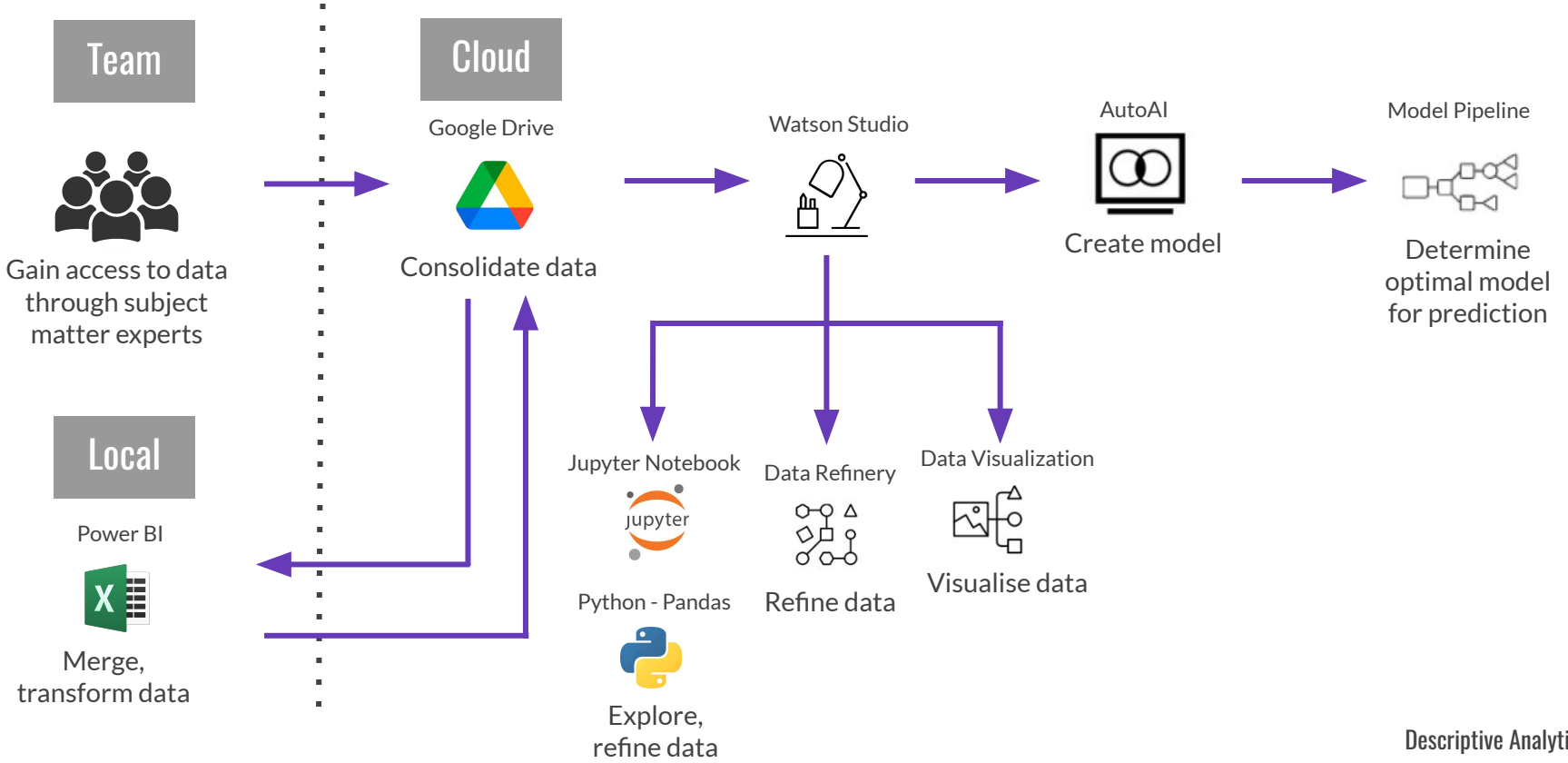
- Characteristics:
  - Must be able to distinguish degree of competence
  - Conceptually correct
  - Question design: distractors (i.e. the 3 wrong answer options) are well-designed to capture common misconceptions / likely mistakes

What data do we have?



Descriptive Analytics

Implementation Architecture

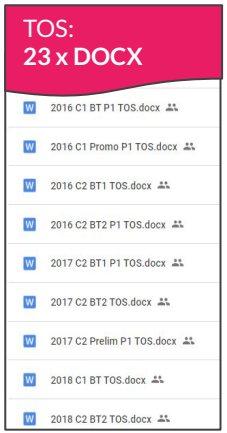


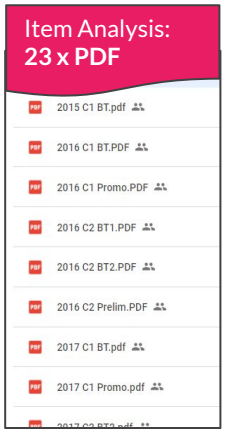
Data Exploration

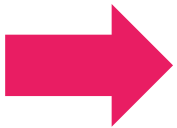
- Pre-Exam
  - Table of specifications for every exam, contains labels of questions
  - By Question
- Post Exam
  - Test statistics (by exam)
  - Item statistics (by questions)

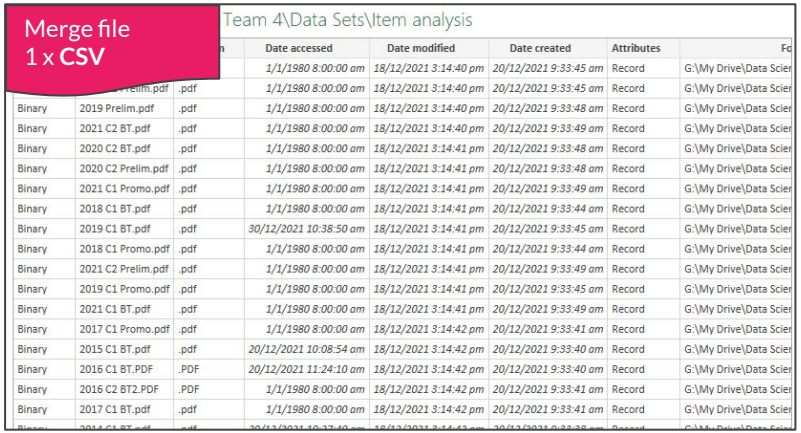
Data Preparation

- Conversion of Word document format to PDF/CSV format
- Merging of two separate datasets into a single document via Power BI
- Merge via common reference column ‘filename’ - e.g. 2016 C1 BT (Year/Level/Test)









			Date accessed	Date modified	Date created	Attributes	Fc
			1/1/1980 8:00:00 am	18/12/2021 3:14:40 pm	20/12/2021 9:33:45 am	Record	G:\My Drive\Data Scier
		.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:40 pm	20/12/2021 9:33:45 am	Record	G:\My Drive\Data Scier
Binary	2019 Prelim.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:40 pm	20/12/2021 9:33:48 am	Record	G:\My Drive\Data Scier
Binary	2021 C2 BT.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:40 pm	20/12/2021 9:33:49 am	Record	G:\My Drive\Data Scier
Binary	2020 C2 BT.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:48 am	Record	G:\My Drive\Data Scier
Binary	2020 C2 Prelim.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:48 am	Record	G:\My Drive\Data Scier
Binary	2021 C1 Promo.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:49 am	Record	G:\My Drive\Data Scier
Binary	2018 C1 BT.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:44 am	Record	G:\My Drive\Data Scier
Binary	2019 C1 BT.pdf	.pdf	30/12/2021 10:38:50 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:45 am	Record	G:\My Drive\Data Scier
Binary	2018 C1 Promo.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:44 am	Record	G:\My Drive\Data Scier
Binary	2021 C2 Prelim.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:49 am	Record	G:\My Drive\Data Scier
Binary	2019 C1 Promo.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:45 am	Record	G:\My Drive\Data Scier
Binary	2021 C1 BT.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:41 pm	20/12/2021 9:33:49 am	Record	G:\My Drive\Data Scier
Binary	2017 C1 Promo.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:42 pm	20/12/2021 9:33:41 am	Record	G:\My Drive\Data Scier
Binary	2015 C1 BT.pdf	.pdf	20/12/2021 10:08:54 am	18/12/2021 3:14:42 pm	20/12/2021 9:33:40 am	Record	G:\My Drive\Data Scier
Binary	2016 C1 BT.PDF	.PDF	20/12/2021 11:24:10 am	18/12/2021 3:14:42 pm	20/12/2021 9:33:40 am	Record	G:\My Drive\Data Scier
Binary	2016 C2 BT2.PDF	.PDF	1/1/1980 8:00:00 am	18/12/2021 3:14:42 pm	20/12/2021 9:33:41 am	Record	G:\My Drive\Data Scier
Binary	2017 C1 BT.pdf	.pdf	1/1/1980 8:00:00 am	18/12/2021 3:14:42 pm	20/12/2021 9:33:41 am	Record	G:\My Drive\Data Scier

What comes out of the Big Merge...

Pre-exam, from Table of Specifications (TOS)
Labeling each question by topics, predicted difficulty, type of question

Post-exam, from Item Analysis
Tells us the exam outcome

Describes the source of test items

475 rows, corresponding to 475 test items across 23 papers

Year of Exam	Name of Exam	Cohort (year of graduation)	Question No.	Main Topic	Assessment Objective - Knowledge with Understanding	Assessment Objective - Handling, applying and Evaluating Information	Predicted difficulty - Easy	Predicted difficulty - Moderate	Predicted difficulty - Difficult	Nature of Question - Calculation [1/0]	Nature of Question - Graphical / Diagram [1/0]	Nature of Question - Qualitative / Conceptual [1/0]	Learning outcome	Graded	Correct	Point Biserial	P-Value
2016	C1 Blk Test	Class of 2017	1	Measurement	0.5	0.5		1		1			1(h)	516	356	0.25	0.690
2016	C1 Blk Test	Class of 2017	2	Measurement	0.5	0.5		1		1			1(i)	516	334	0.32	0.647
2016	C1 Blk Test	Class of 2017	3	Kinematics	1				1			1	2(h)	516	227	0.21	0.440
2016	C1 Blk Test	Class of 2017	4	Kinematics		1			1		1		2(b), 3(e)	516	345	0.37	0.669
2016	C1 Blk Test	Class of 2017	5	Dynamics		1		1		1			3(f)	516	434	0.29	0.841
2016	C1 Blk Test	Class of 2017	6	Dynamics	1		1					1	3(c), 3(f)	516	266	0.45	0.516
2016	C1 Blk Test	Class of 2017	7	Dynamics	1			1				1	3(h)	516	307	0.38	0.595
2016	C1 Blk Test	Class of 2017	8	Forces		1		1				1	4(h)	516	295	0.36	0.572
2016	C1 Blk Test	Class of 2017	9	Forces		1		1			1		4(n)	516	412	0.39	0.798
2016	C1 Blk Test	Class of 2017	10	Work, Energy and Power	1			1			0.5	0.5	4(b),5a)	516	385	0.43	0.746
2016	C1 Blk Test	Class of 2017	11	Work, Energy and Power		1			1			1	3(f),5(c)	516	444	0.29	0.860
2016	C1 Blk Test	Class of 2017	12	Motion in a Circle		1			1	1			6(f)	516	166	0.46	0.322
2016	C1 Blk Test	Class of 2017	13	Motion in a Circle		1		1		1			6(e), 1(k)	516	196	0.52	0.380
2016	C1 Blk Test	Class of 2017	14	Gravitation	1		1			0.5		0.5	7(b)	516	318	0.43	0.616

Dataset: Table of Specifications and Item Analysis

Attribute	Description	Type	Count
File Name	File name convention used for each paper	String	23 papers
Year of Exam	Year of when exam paper was conducted	Integer	6 years
Date of Exam	Date of when exam paper was conducted	String	23 dates
Name of Exam	Name of exam paper	String	23 papers
Cohort (Year of Graduation)	Graduation year for the respective cohort	Integer	6 years
Question	Question number for respective exam	Integer	475 questions
Topic Acronym	Acronyms used for Physics topic names	String	19 topics
Assessment Objective - Knowledge with Understanding [1/0]	Assessment objective #1 for specific exam question	Decimal	164 entries
Assessment Objective - Handling, Applying and Evaluating Information [1/0]	Assessment objective #2 for specific exam question	Decimal	320 entries

Dataset: Table of Specifications and Item Analysis

Attribute	Description	Type	Count
Predicted Difficulty - Easy	Level of difficulty that is perceived for specific exam question	Decimal	137 entries
Predicted Difficulty - Moderate		Decimal	271 entries
Predicted Difficulty - Difficult		Decimal	75 entries
Nature of Question - Calculation [1/0]	Nature of question #1 for specific exam question	Decimal	255 entries
Nature of Question - Graphical / Diagram [1/0]	Nature of question #2 for specific exam question	Decimal	139 entries
Nature of Question - Qualitative / Conceptual [1/0]	Nature of question #3 for specific exam question	Decimal	183 entries
Learning Outcome 1	Possible learning outcome determined for specific exam question	String	315 entries
Remarks	Remarks relevant to specific exam question	String	144 entries

Data Exploration

```
df.describe()
```

	Year of Exam	Question	Assessment Objective - Knowledge with Understanding [1/0]	Assessment Objective - Handling, applying and Evaluating Information [1/0]	Predicted difficulty - Easy	Predicted difficulty - Moderate	Predicted difficulty - Difficult	Difficulty	Nature of Question - Calculation [1/0]	Nature of Question - Graphical /Diagram [1/0]	Nature of Question - Qualitative / Conceptual [1/0]	Graded	Correct	Point Biserial	P-Value
count	475.000000	475.000000	164.000000	320.000000	137.000000	271.000000	75.000000	475.000000	255.000000	139.000000	183.000000	475.000000	475.000000	475.000000	475.000000
mean	2018.235789	12.132632	0.975610	0.981250	0.992701	0.992620	0.973333	1.858947	0.858824	0.748201	0.836066	504.244211	317.983158	0.349811	0.631078
std	1.769164	8.205772	0.108034	0.110393	0.085436	0.074168	0.215816	0.642302	0.234081	0.250898	0.246759	22.585593	93.191761	0.091133	0.184376
min	2016.000000	1.000000	0.500000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.500000	0.000000	466.000000	50.000000	-0.020000	0.101000
25%	2016.500000	6.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.500000	0.500000	0.500000	488.000000	252.500000	0.290000	0.502000
50%	2018.000000	11.000000	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000	1.000000	0.500000	1.000000	497.000000	329.000000	0.360000	0.659000
75%	2020.000000	16.500000	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000	1.000000	1.000000	1.000000	512.000000	390.000000	0.420000	0.778000
max	2021.000000	40.000000	1.000000	1.000000	1.000000	1.000000	2.000000	3.000000	1.000000	1.000000	1.000000	556.000000	523.000000	0.560000	0.957000

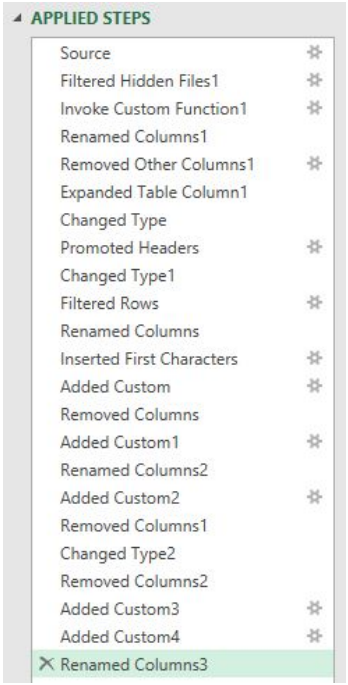
Data Cleansing / Transformation

Within PowerBI:

- Align on appropriate headers for data [Cohort/Topic/Difficulty/Nature of Qn]
- Remove unnecessary columns [e.g. Filename, date, question number]

Within Excel:

- Further formatting of data type (numbers, date, string)
- Data transformation of difficulty from Easy/Moderate/Difficult to ranking system
- Data cleansing to arrive at standard list of Physics topics



Topic	I Easy(E)	II Moderate (C)	III Difficult (A)	Difficulty
Kinematics+Dynamics			1	3
Dynamics		1		2
Dynamics	1			1

Data Transformation

Acronym	Names of Physics topic and sub-topics
AC	Alternating Current
CM	Motion in a Circle
DC_COE	DC/Current of Electricity, Current of Electricity
Dynamics	Dynamics
E_field	Electric Fields, Electric Potential
EM	Electromagnetism, B Field of Solenoid
EMI	Electromagnetic Induction, Magnetic Flux Linkage, Faraday's Law
Forces	Forces, Torque
WEP	Work, Energy, Power

Acronym	Names of Physics topic and sub-topics
Gravitation	Gravitation, G-force, Gravitational Field
Kinematics	Kinematics
Measurement	Measurement, Vectors
Nuclear	Nuclear Physics
Quantum	Quantum Physics, Photoelectric, Modern Physics, Line Spectra
SHM	Oscillations, Max Speed (SHM), Energy (SHM)
Superposition	Superposition, Standing Waves, Single Slit
Thermal	Thermal Physics, Temperature & Ideal Gases, First Law of Thermodynamics, Ideal Gases
Waves	Wave Motion, Phase, Polarisation

Data Transformation

```
dummy_variable_1 = pd.get_dummies(df['Topic Acronym'])
dummy_variable_1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 473 entries, 0 to 472
Data columns (total 18 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   AC             473 non-null    uint8
 1   CM             473 non-null    uint8
 2   DC_COE         473 non-null    uint8
 3   Dynamics       473 non-null    uint8
 4   EM             473 non-null    uint8
 5   EMI            473 non-null    uint8
 6   E_Field        473 non-null    uint8
 7   Forces         473 non-null    uint8
 8   Gravitation    473 non-null    uint8
 9   Kinematics     473 non-null    uint8
 10  Measurement    473 non-null    uint8
 11  Nuclear        473 non-null    uint8
 12  Quantum        473 non-null    uint8
 13  SHM            473 non-null    uint8
 14  Superposition  473 non-null    uint8
 15  Thermal        473 non-null    uint8
 16  WEP            473 non-null    uint8
 17  Waves          473 non-null    uint8
dtypes: uint8(18)
memory usage: 8.4 KB
```

- In preparation to generate AutoAI models, we also performed One-Hot Encoding of topic acronyms
- Numerical values were assigned to the 18 physics topics using Python

Data Transformation

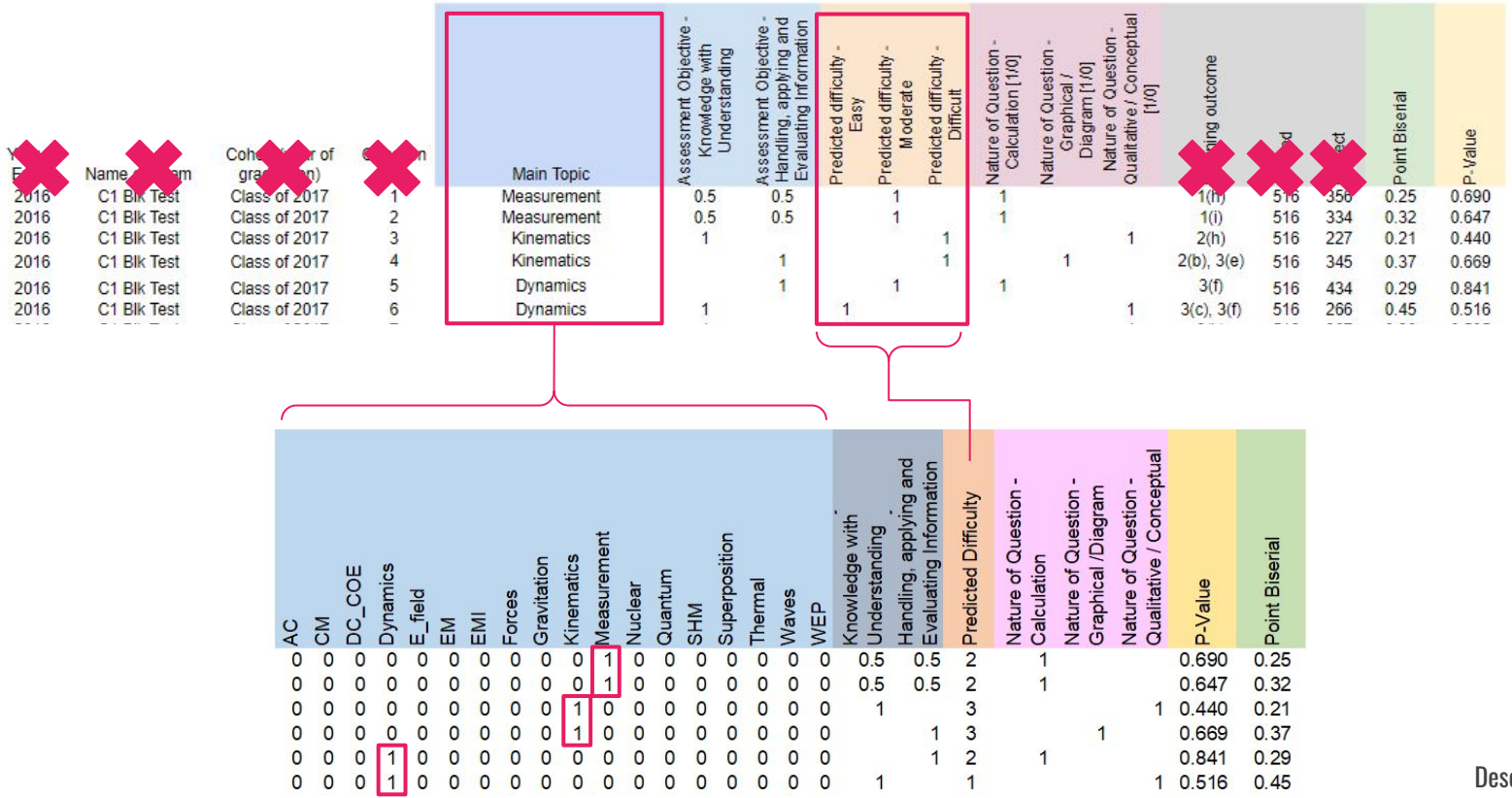
- Merged the indicator variables to the main dataframe using pandas' concat method
- Removed the column 'Topic Acronym'

```
df2 = pd.concat([df, dummy_variable_1], axis=1)
df2.drop("Topic Acronym", axis = 1, inplace=True)
```

```
df2.head()
```

Question	Assessment Objective - Knowledge with Understanding [1/0]	Assessment Objective - Handling, applying and Evaluating Information [1/0]	Predicted difficulty - Easy	Predicted difficulty - Moderate	...	Gravitation	Kinematics	Measurement	Nuclear	Quantum	SHM	Superposition	Thermal	WEP	Waves
1	0.5	0.5	NaN	1.0	...	0	0	1	0	0	0	0	0	0	0
2	0.5	0.5	NaN	1.0	...	0	0	1	0	0	0	0	0	0	0
3	1.0	NaN	NaN	NaN	...	0	1	0	0	0	0	0	0	0	0

Overview of Transformation



Name of Exam	Cohort (year of graduation)	Question No.	Main Topic	Assessment Objective - Knowledge with Understanding	Assessment Objective - Handling, applying and Evaluating Information	Predicted difficulty - Easy	Predicted difficulty - Moderate	Predicted difficulty - Difficult	Nature of Question - Calculation [1/0]	Nature of Question - Graphical / Diagram [1/0]	Nature of Question - Qualitative / Conceptual [1/0]	Point Biserial	P-Value
C1 Blk Test	Class of 2017	1	Measurement	0.5	0.5		1		1			0.25	0.690
C1 Blk Test	Class of 2017	2	Measurement	0.5	0.5		1		1			0.32	0.647
C1 Blk Test	Class of 2017	3	Kinematics	1				1			1	0.21	0.440
C1 Blk Test	Class of 2017	4	Kinematics		1			1		1		0.37	0.669
C1 Blk Test	Class of 2017	5	Dynamics		1		1		1			0.29	0.841
C1 Blk Test	Class of 2017	6	Dynamics	1		1					1	0.45	0.516

AC	CM	DC_COE	Dynamics	E_field	EM	EMI	Forces	Gravitation	Kinematics	Measurement	Nuclear	Quantum	SHM	Superposition	Thermal	Waves	WEP	Knowledge with Understanding	Handling, applying and Evaluating Information	Predicted Difficulty	Nature of Question - Calculation	Nature of Question - Graphical /Diagram	Nature of Question - Qualitative / Conceptual	P-Value	Point Biserial
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.5	0.5	2	1			0.690	0.25
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.5	0.5	2	1			0.647	0.32
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1		3			1	0.440	0.21
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0		1	3		1		0.669	0.37
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		1	2	1			0.841	0.29
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		1			1	0.516	0.45

Data Validation

- Data quality assurance
  - Created a formula to flag missing values as ‘FALSE’
  - Exclude rows with missing/incomplete data
  - Run validation check on finalised data attributes (e.g. topic subjects)
  - Check that data attributes are correctly identified (String/Integer/Decimal/Date)

Diagnostic Analytics

Diagnostics

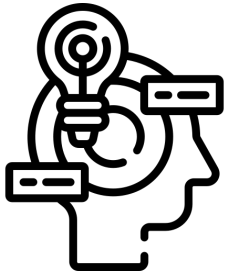
Big picture view

- Understand the metrics
- Sense-making, relating back to business understanding

Deep Dive

- Outcomes* vs Topics
- Outcomes vs Type of questions

**Outcomes: p-value, point-biserial*



End Goal

- Develop insights
- Angle to focus on for our predictive analytics

Recall...

Pre-exam labeling: Topic, Predicted Difficulty, Nature of Question

Post-exam metrics: P-value, point biserial

475 test items across 23 papers

Main Topic	Assessment Objective - Knowledge with Understanding	Assessment Objective - Handling, applying and Evaluating Information	Predicted difficulty - Easy	Predicted difficulty - Moderate	Predicted difficulty - Difficult	Nature of Question - Calculation [1/0]	Nature of Question - Graphical / Diagram [1/0]	Nature of Question - Qualitative / Conceptual [1/0]	Point Biserial	P-Value
Measurement	0.5	0.5		1		1			0.25	0.690
Measurement	0.5	0.5		1		1			0.32	0.647
Kinematics	1				1			1	0.21	0.440
Kinematics		1			1		1		0.37	0.669
Dynamics		1		1		1			0.29	0.841
Dynamics	1		1					1	0.45	0.516

P-Value and Point Biserial

P-Value

Measure of: Difficulty

Proportion of candidates who get test item correct

>0.85	Very easy, not discriminating
0.7 - 0.85	Easy
0.5 - 0.7	**Moderate; Ideal range**
0.25 - 0.5	Difficult
< 0.25	Very difficult, not discriminating

Point Biserial

Measure of: Reliability and Discrimination

Correlation of

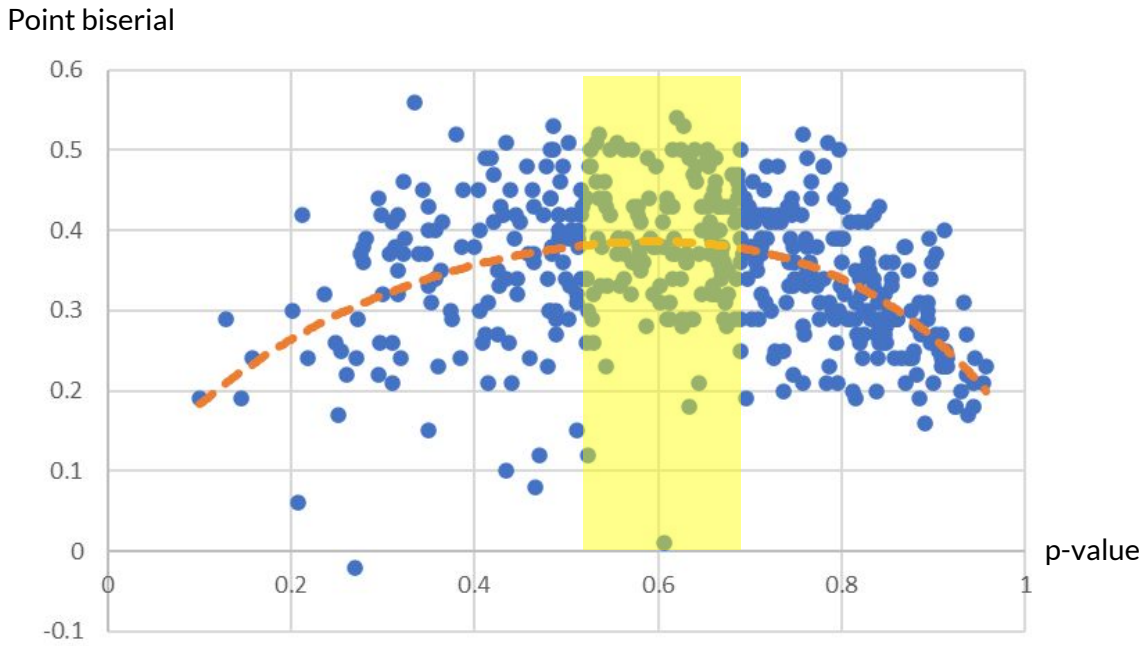
- candidates' overall score less that test item (eg, 24 out of 29) *vs*
- respective candidates' marks for that item (1 or 0).

Negative	Big problem; may indicate major error
Around 0	No relationship between question score and assessment score; need to review what's wrong with item
0.05 - 0.2	Poor discrimination
0.21 - 0.3	So-so
0.31 - 0.45	Quite discriminating, served purpose.
>0.45	Very discriminating, good as a model

Source: https://www.researchgate.net/publication/315321558_Using_reliability_and_item_analysis_to_evaluate_a_teacher-developed_test_in_educational_measurement_and_evaluation

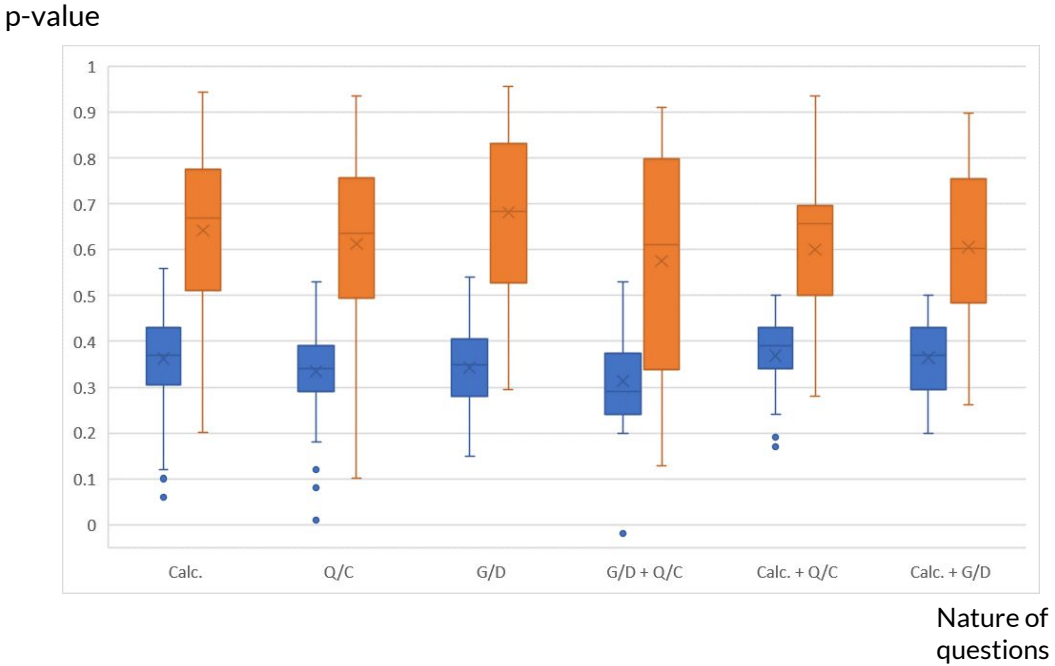
Overview: Point Biserial vs P-value

- Scatter plot - all 475 test items
- Why these two metrics?
- Question too easy (p-value high), everyone also wrong, point biserial low; too difficult, everyone also wrong, point biserial also low
- Good question should aim for band of 0.5-0.7



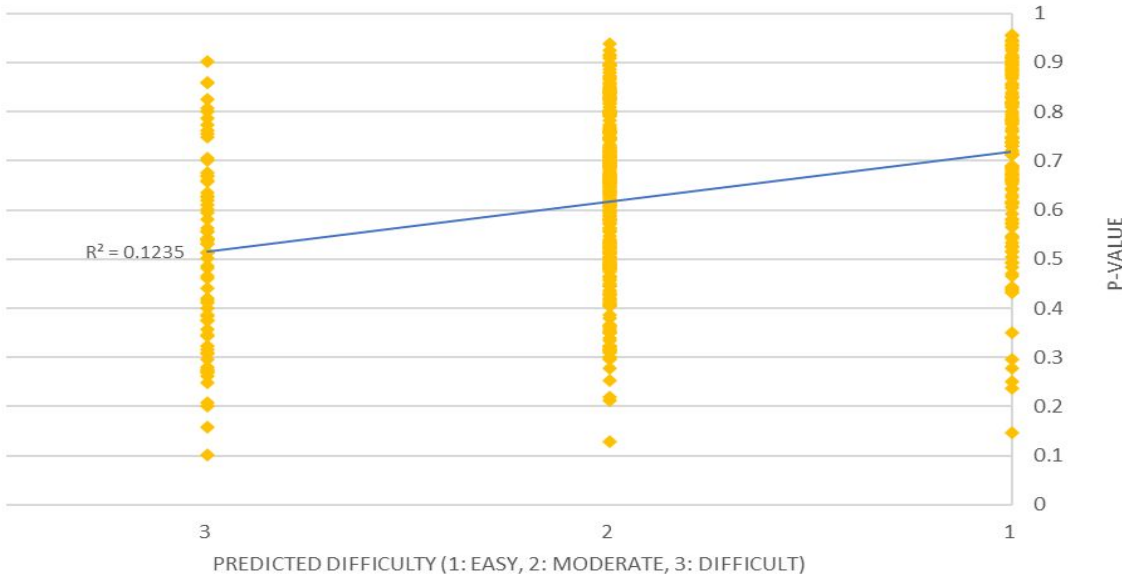
Deep Dive: P-value vs Nature of questions

- Box and whisker plot
  - Box: 25th to 75th percentiles of test items
  - Cross: Mean
  - Whisker: Max, Min
- Varied spread, but similar means
- May not be too good an angle; categorisations are narrow; also less helpful from use case point of view



Deep Dive: P-value vs Predicted Difficulty

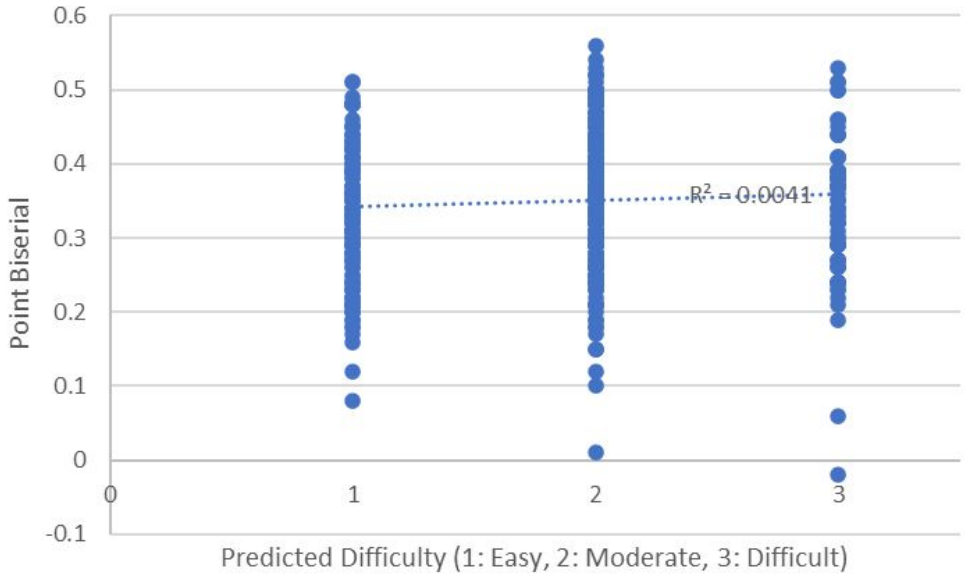
- Scatterplot
- $R^2 = 0.12 \Rightarrow$ weak positive correlation
- Setters are not too good at predicting the difficulty of the questions they set!



>0.85	1
0.7 - 0.85	2
0.5 - 0.7	3
0.25 - 0.5	4
< 0.25	5

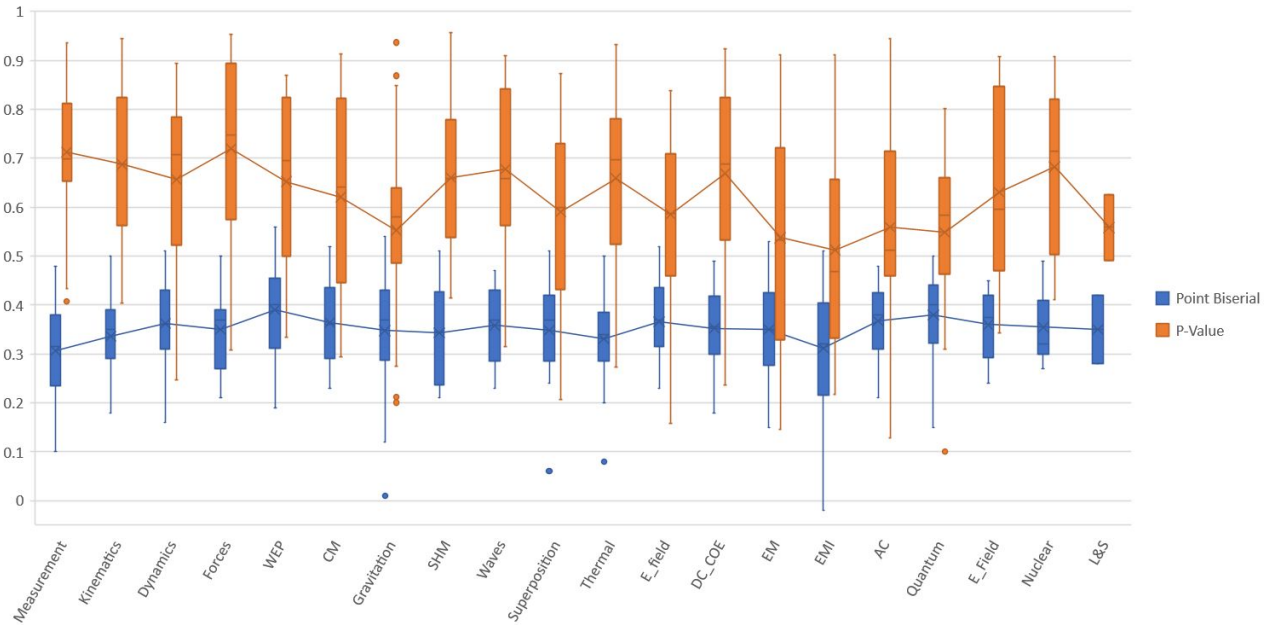
Deep Dive: Point Biserial vs Predicted Difficulty

- $R^2 = 0.004$
- Predicted difficulty has no correlation with point-biserial



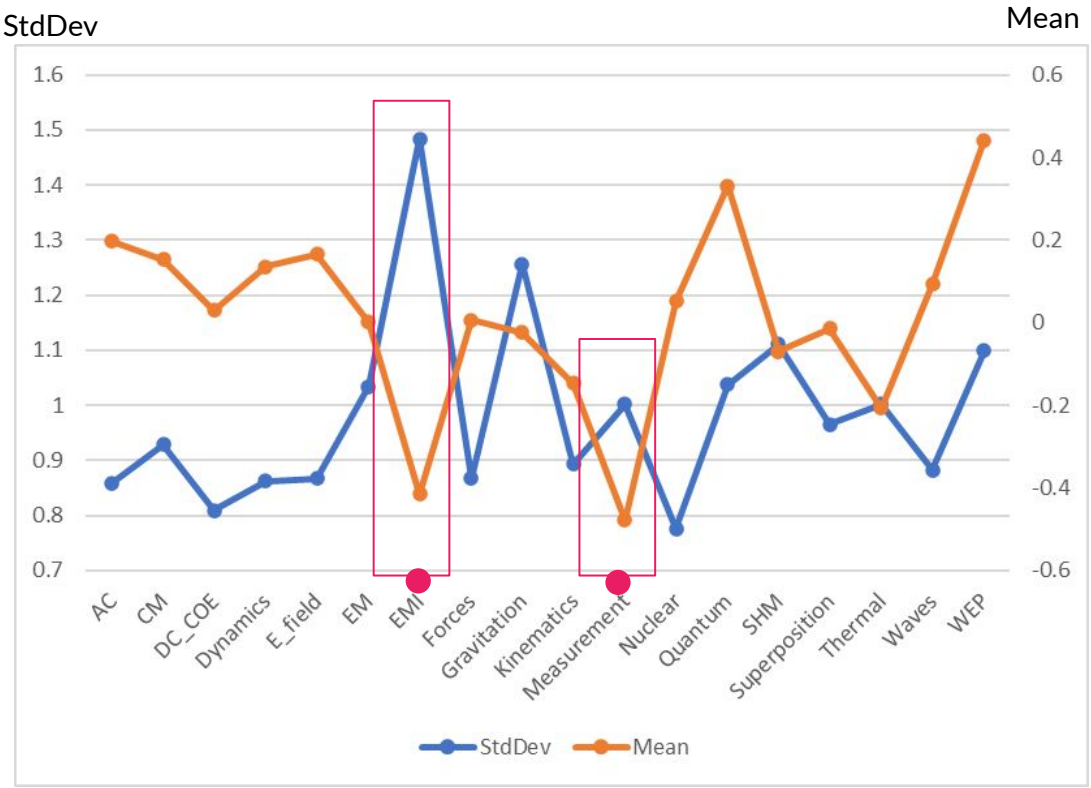
Deep Dive: P-value and Point Biserial vs Topics

- Getting a feel of how of the variation across topics
- *It's a jungle out there!*
- Absolute values not so good to compare. Next step: normalise



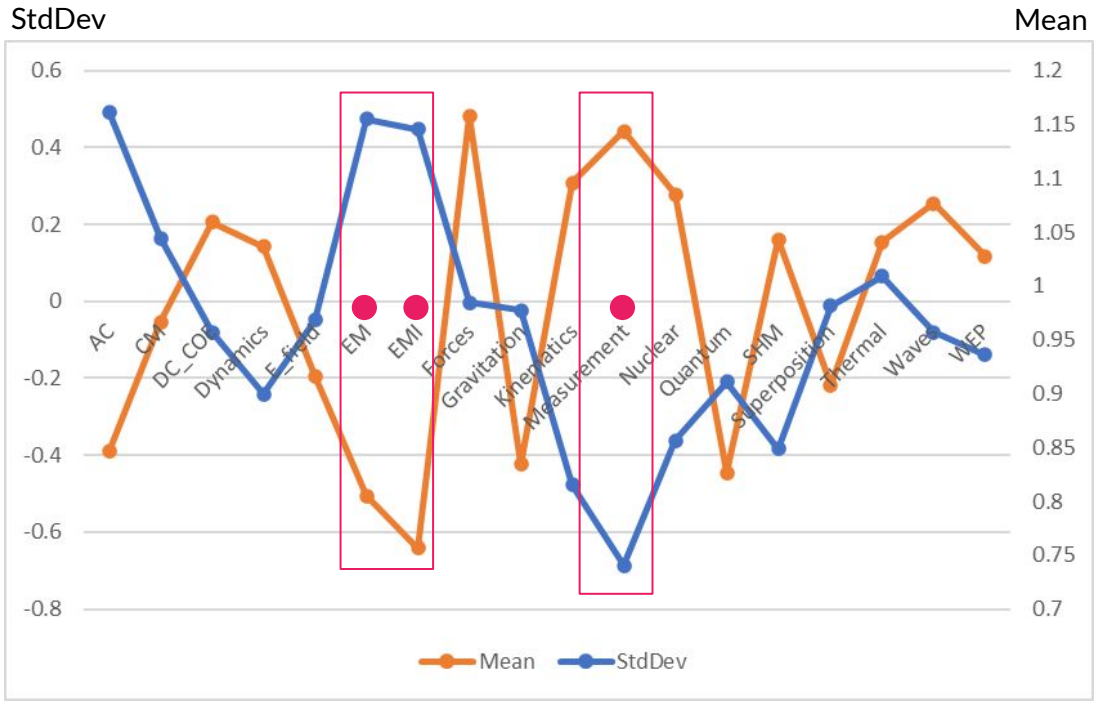
Normalised P-values across Topics

- Standard deviation tells us the spread of questions in terms of difficulty
- Mean tells us on average, are questions of this topic difficult
- Difficult test items:
  - EMI - large spread
  - Measurement - small spread



Normalised Point Biserial across Topics

- Standard deviation tells us the spread of questions in terms of how well they discriminate
- Mean tells us on average, are questions of this topic discriminating
- Topics of concern:
  - EM, EMI
- Doing well:
  - Measurement



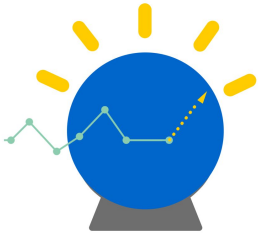
Insights from Diagnostics

Study by topics alone most meaningful, due to greater diversity

Predicted Difficulty > Nature of Question

P-value shows overall more correlation to various factors

Our data is to likely to provide better predictive power in p-value



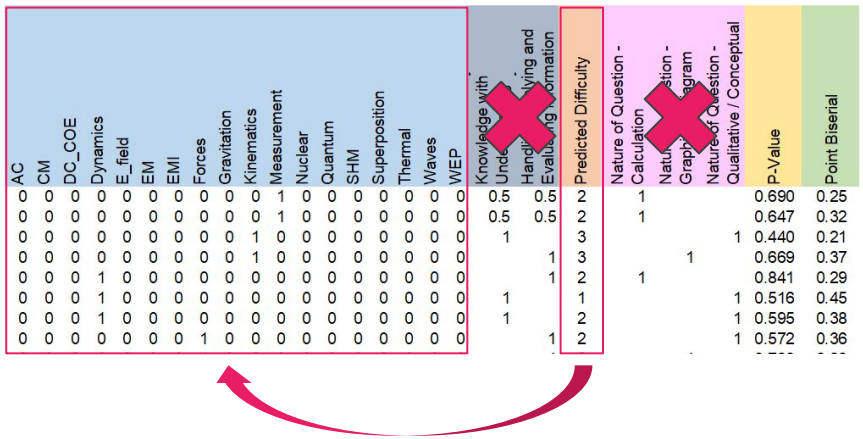
Predict: P-value

Primary angle: Topic; Secondary angle: Predicted Difficulty

Further date transformation

Predictive Analytics

What data we eventually run AutoAI on…



AC	CM	DC_COE	Dynamics	E_field	EM	EMI	Forces	Gravitation	Kinematics	Measurement	Nuclear	Quantum	SHM	Superposition	Thermal	Waves	WEP	Knowledge with Unde	Handl ying and Evalu ng_ormation	Predicted Difficulty	Nature of Question - Calculation	Natu stion - Graph gram	Nature of Question - Qualitative / Conceptual	P-Value	Point Biserial
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.5	0.5	2	1			0.690	0.25
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.5	0.5	2	1			0.647	0.32
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1		3			1	0.440	0.21
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0		1	3		1		0.669	0.37
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		1	2	1			0.841	0.29
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		1			1	0.516	0.45
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		2			1	0.595	0.38
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0		1	2			1	0.572	0.36

AC	CM	DC_COE	Dynamics	E_field	EM	EMI	Forces	Gravitation	Kinematics	Measurement	Nuclear	Quantum	SHM	Superposition	Thermal	Waves	WEP	Point Biserial	Point Biserial
0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0.25	0.25
0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0.32	0.32
0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0.21	0.21
0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0.37	0.37
0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.29	0.29
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.45	0.45
0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.38	0.38

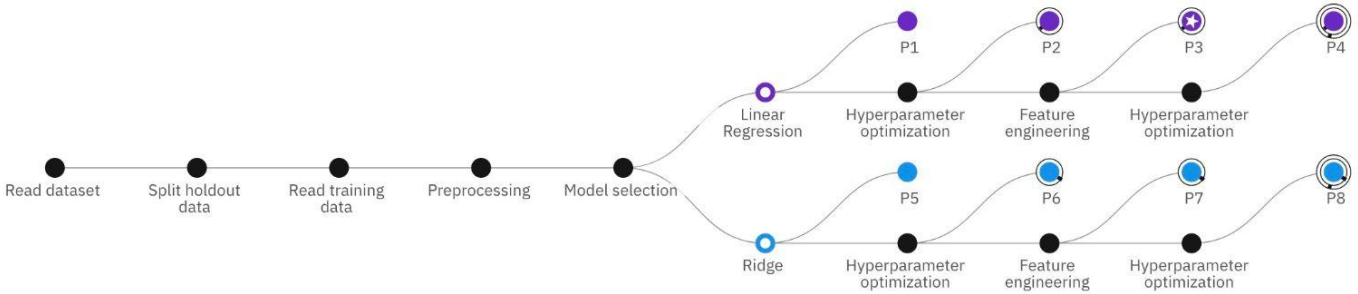
What data we eventually run AutoAI on…

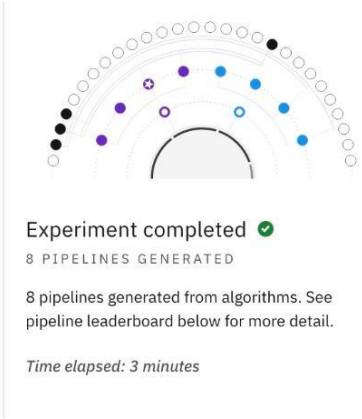
P-Value	AC	CM	DC_COE	Dynamics	E_field	EM	EMI	Forces	Gravitation	Kinematics	Measurement	Nuclear	Quantum	SHM	Superposition	Thermal	Waves	WEP
0.69	0	0	0	0	0	0	0	0	0	0	1.38	0	0	0	0	0	0	0
0.647	0	0	0	0	0	0	0	0	0	0	1.294	0	0	0	0	0	0	0
0.44	0	0	0	0	0	0	0	0	0	1.32	0	0	0	0	0	0	0	0
0.669	0	0	0	0	0	0	0	0	0	2.007	0	0	0	0	0	0	0	0
0.841	0	0	0	1.682	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.516	0	0	0	0.516	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.595	0	0	0	1.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.572	0	0	0	0	0	0	0	1.144	0	0	0	0	0	0	0	0	0	0
0.798	0	0	0	0	0	0	0	1.596	0	0	0	0	0	0	0	0	0	0
0.746	0	0	0	0	0	0	0	1.492	0	0	0	0	0	0	0	0	0	0
0.86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.58
0.322	0	0.966	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.38	0	0.76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.616	0	0	0	0	0	0	0	0	0.616	0	0	0	0	0	0	0	0	0
0.61	0	0	0	0	0	0	0	0	1.22	0	0	0	0	0	0	0	0	0
0.93	0	0	0	0	0	0	0	0	0	0	0.93	0	0	0	0	0	0	0

Training Data Models with IBM Watson Studio's AutoAI

Progress Map

Target Variable: P-Value





Use of AutoAI toolkit in IBM Watson Studio to:

- Automatically prepare data - 90:10 split ratio of training and holdout data
- Apply machine learning algorithms
- Build model pipelines that best suit our dataset and business case

Training Data Models with IBM Watson Studio's AutoAI

Pipeline Leaderboard: Top 4

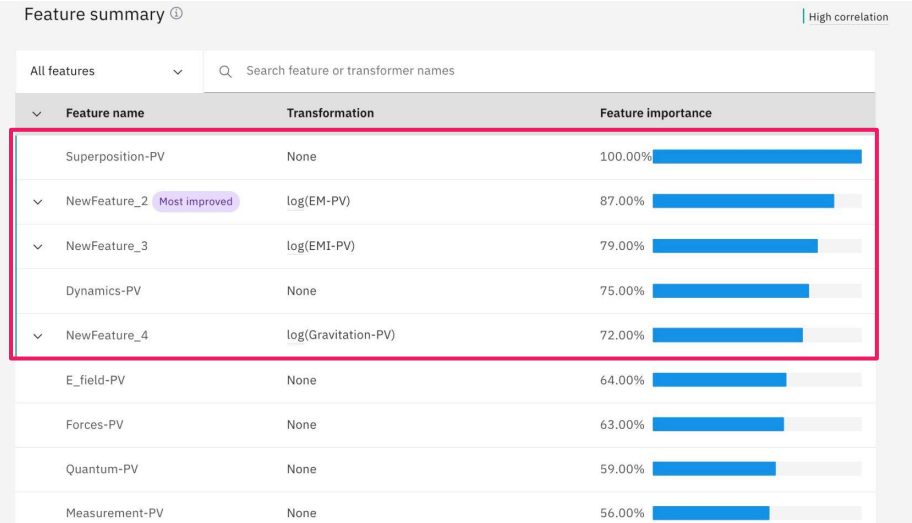
	Rank ↑	Name	Algorithm	RMSE (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 3	Linear Regression	0.163	HPO-1 FE	00:00:23
	2	Pipeline 4	Linear Regression	0.163	HPO-1 FE HPO-2	00:00:03
	3	Pipeline 8	Ridge	0.163	HPO-1 FE HPO-2	00:00:21
	4	Pipeline 7	Ridge	0.163	HPO-1 FE	00:00:36

Determined linear regression to be most basic model to start with:

- Dealing with numerical values
- Values are linearly related to each other
- Function is multiple linear regression

Training Data Models with IBM Watson Studio's AutoAI

Ranked #1: Pipeline 3 - (Multiple) Linear Regression



Model evaluation

Model evaluation measure

Measures	Holdout score	Cross validation score
Root mean squared error	0.157	0.163
R squared	0.144	0.218
Explained variance	0.146	0.226
Mean squared error	0.025	0.027
Mean squared log error	0.010	0.011
Mean absolute error	0.124	0.131
Median absolute error	0.104	0.109
Root mean squared log error	0.098	0.103

Considering that the dataset is restrictive, the model's performance is deemed to be satisfactory

Multiple Linear Regression Equation: $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p$

Source: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704-ep713_multivariablemethods/bs704-ep713_multivariablemethods2.html

Validation of Data Model

P-Value Threshold Ranges:

>0.85	Very easy, not discriminating
0.7 - 0.85	Easy
0.5 - 0.7	Moderate; Ideal range
0.25 - 0.5	Difficult
< 0.25	Very difficult, not discriminating

Results from Testing Model:

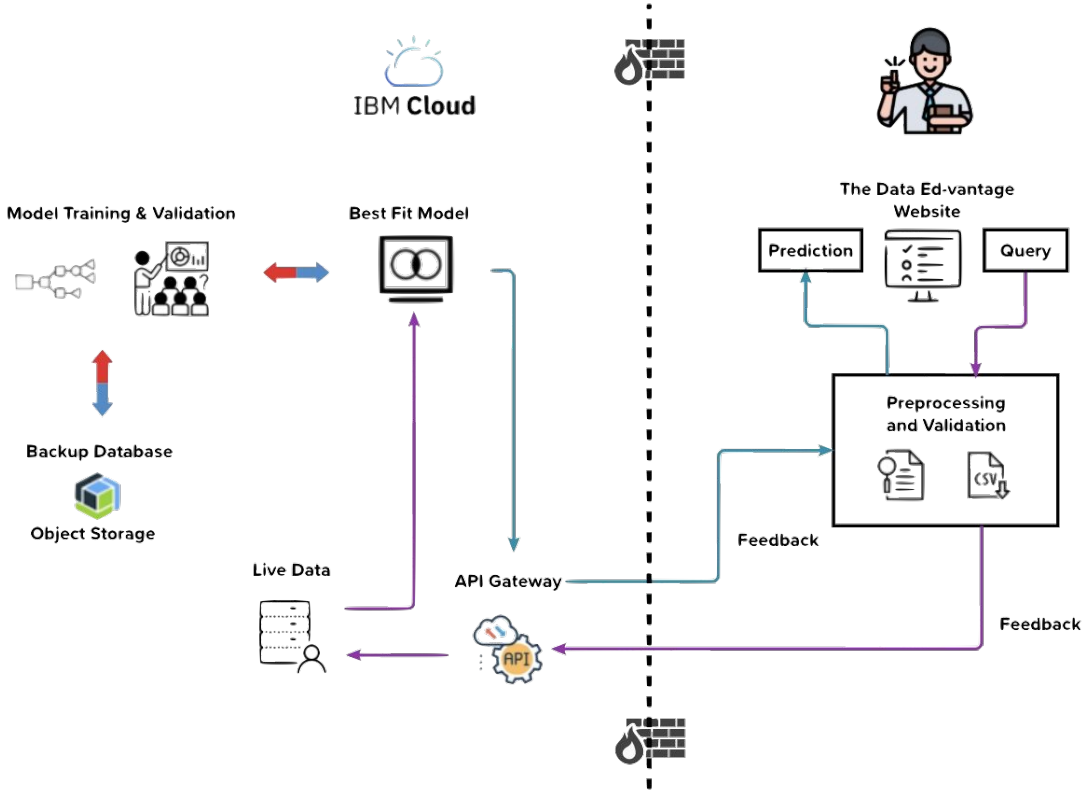

```
{
    "predictions": [
        {
            "fields": [
                "prediction"
            ],
            "values": [
                [
                    0.5651962757110596
                ],
                [
                    0.5269246101379395
                ],
                [
                    0.535916268825531
                ],
                [
                    0.5866577625274658
                ],
                [
                    0.30093348026275635
                ]
            ]
        }
    ]
}
```


	By Subject Matter Experts		By Model	
Feature (i.e. Topic)	**Level of Difficulty**	**P-Value**	**Predicted P-Value**	**Predicted Level of Difficulty**
Superposition	3 (Difficult)	0.271	0.565	2 (Moderate)
EMI	3 (Difficult)	0.269	0.527	2 (Moderate)
EM	3 (Difficult)	0.316	0.536	2 (Moderate)
Dynamics	3 (Difficult)	0.248	0.587	2 (Moderate)
Gravitation	3 (Difficult)	0.201	0.301	3 (Difficult)

↑ Level of Difficulty = ↓ P-Value = ↑ Quality of Exam Question

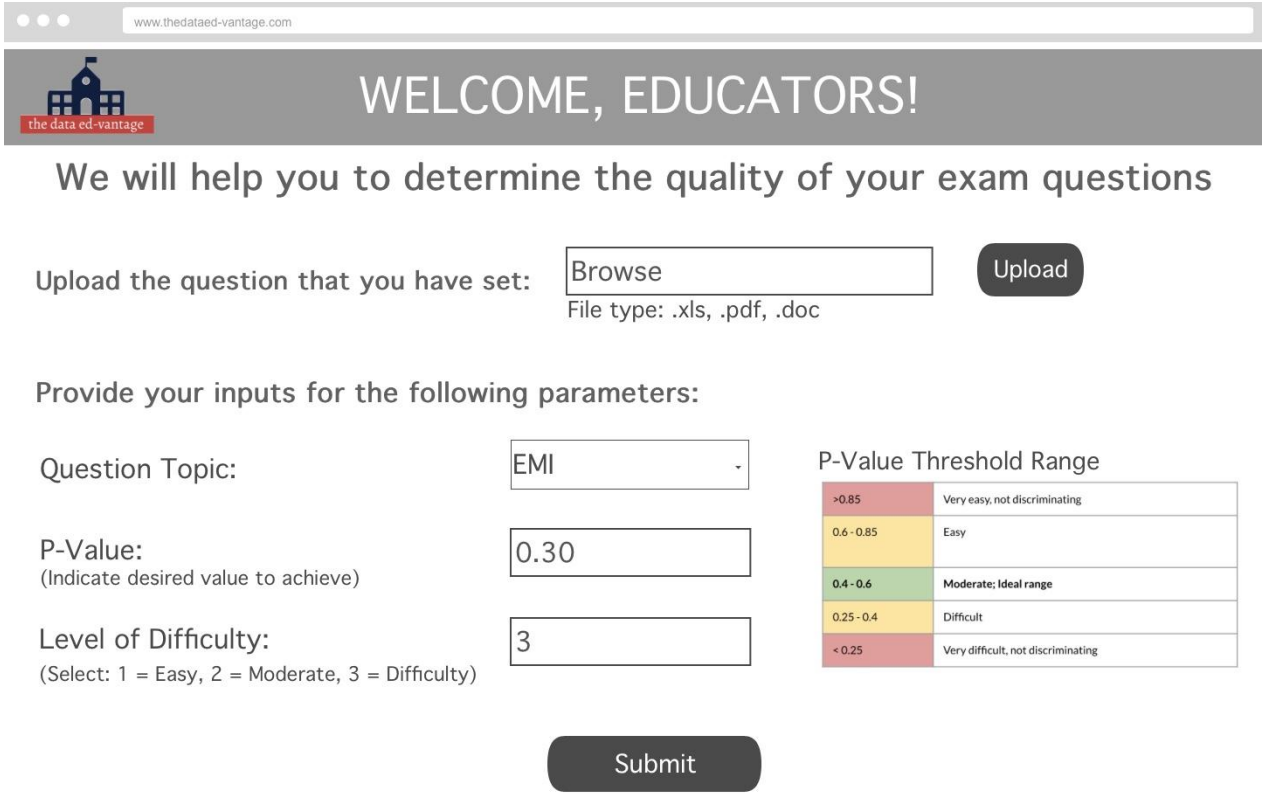
Prescriptive Analytics

Model Deployment

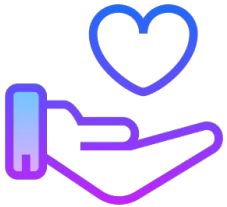


- Web-based interface that focuses on user's experience - convenient, hassle free and concise
- Automation of data pre-processing based on user's inputs
- Zero downtime at all times
- Consistent monitoring and managing of current model using DevSecOps methodology

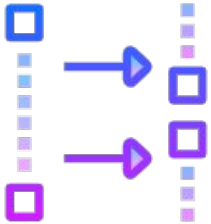
Website Interface - Wireframe



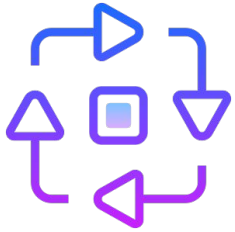
Environment Feedback



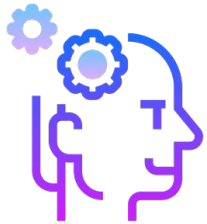
User-friendliness of website



Periodic re-fitting of model triggered by changes in data attributes



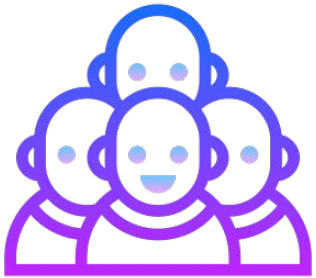
Continuous refining of model to reduce RMSE of prediction accuracy



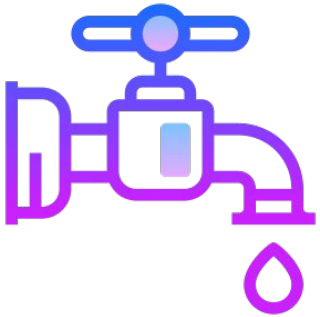
Subject matter experts' guidance on how they want to set the exam questions, write conditions

Summary & Reflections

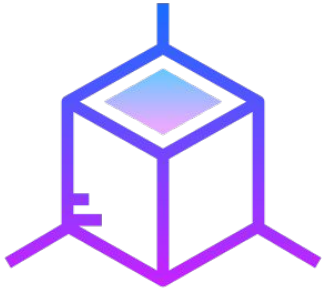
Constraints



Insufficient variant of datasets due to students competency level being too similar



Difficulty in getting more quantity of data

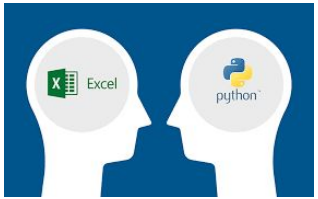


With single dataset, limited fields result in limited angles to analyze data

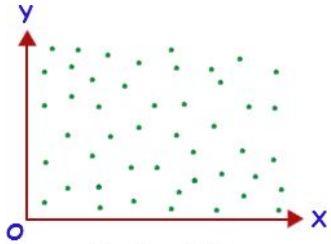
Challenges

Consolidation of data from files of differing formats

Limitations of CUH (capacity unit hour) available for Machine Learning in IBM Watson Studio Lite plan



Adapting to using data analytics tools (PowerBI, Python, IBM Watson) instead of reverting to manual data transformation



Finding trends/correlation between attributes and limited fields

Conclusion

Design thinking plays a crucial role in enabling all parties to align focus and keep our target users in addressing their pain points

Beneficial to possess knowledge and skills in using tools that enable a more efficient way of handling huge datasets

Machine learning serves to make visible the gap between expectations of teachers vs students' actual level of performance for specific topics

Future Plan/Enhancement/Consideration

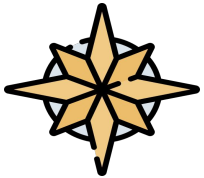
Possible work directions:

- Use of dummy datasets (through data manipulation) to simulate different scenarios to test the model's elasticity using extreme values and improve the model's understanding of our data
- Additional data preprocessing required to achieve consistent formatting of data obtained from other junior colleges
- Further exploration of more complex algorithms for our data to achieve greater accuracy of results

Learning Takeaways

"If we lack domain knowledge, it is best to first start with looking at the data." - Loretta

"Must learn to love 'Data' like a commodity for its intrinsic value." - Siang Khim



"Business understanding is the north-star in the midst of chaos." - Wee Tong

"The aim is to turn data into information, and information into insight." - Melodie

"Teamwork is very important." - Faz

Thank You