EXAM AUTOMATION

" We are not building an automation for existing flawed system , We are proposing a strict system which can be automated that will be beneficial for the students and the teachers in the long run .... "

The goal is to automate the paper correction process even for lengthy answers for different subjects. The answers for any questions that are asked frequently in the examinations of universities or institution or even schools can be categorized into the following:

1. Cognitive
2. Retentive
3. Definitive

Types of Answers:

Cognitive: Humane answers will have the students creative touch and interpretations of the issues/topics asked in the question. It will be biased to the opinion of the student itself. These kinds of answers will mostly occur in subjects such as Linguistics.

Questions such as Debate. Poem or Lesson interpretation. Reasoning etc.

Given the fact that there are advanced techniques that to get these done, but the amount of effort both in data annotation and computation is huge and does not fit in the data industry.

These type answers are good to be set for manual review and it is very important for a teacher to manually go through the answers of each student and understand the level of creativity that they put in to make the answers more beautiful and effective.

Retentive: Retentive answers are related to subjects which are theoretical and do not come under the following type of textual representations:

- Do not have mathematical notations.
- Do not any derivations such as in mathematics and physics or even some parts of chemistry.

Answers that are definitions of certain process, certain names and nouns etc. Answers that are explanation to certain phenomenon which can be matched with an **ANSWER KEY**" are called Cognitive.

Subjects such as Biology, Zoology and any theoretical subjects whose answers have such characteristics. These can be automated.

Definitive: These questions have definite one word or numeric answer which can be chosen from Multiple Choice Questions. These are already automated and can be scaled up more in the system.

Mathematics, Physics, Chemistry and other theoretical subjects where a question can be fit into MCQ format will come under this section.

Best Practices:

- Question such as giving an example can be transferred to MCQ
- Questions that have diagrammatic needs can be transferred to MCQ in the form label detection which again can be put into an MCQ format.
- Questions that have any formula notation can also be transferred to the MCQ section as the memory power of the student is test in all the above-mentioned scenarios

How Robust the model is going to be? How to measure Robustness?

Let's start by understanding the data collection and gradually move to measuring part.

Characteristics of answer-questions:

- Definitions
- Short answers
- Not more than 5 sentences
- Not more than 50 - 80 words.

What will the Target Variable?

Relevancy Score: This is the score assigned by the teacher for each answer based upon their own intuition and the expected answer. They will check for the relevancy of the answer with the question and match it with answer key if the important points are covered.

- Important points to be noted:
- This score will be dependent on the key as well as the answer as the teacher gives the score based on these two sources only i.e. Unbiased Scoring
- The features that must be extracted should consider both the key and the answer and will predict the relevancy score.
- The relevancy score range (0 - 5)
- After the prediction of the relevancy score for a given category of a question, we can essentially convert this into a mark segment (2 marks, 5 marks, 10 marks etc.)

For each question there must be minimum 200 answers with annotated relevancy score.

Data collection Steps:

1. Decide the subjects for which you want to collect the answers
2. Sit with the respective professors to understand the intuition of how their mind works while scoring an answer to a question based upon the given key
3. Explain the concept of Retentive questions to them. Using their guidance for each chapter in the subject frame retentive questions.
4. These intuitions will differ from subject to subject. Terminology intensive subjects such as biology, botany, geography will have different distributions of features in comparison with subjects such as History, civics or economics even some parts of chemistry. Hence subject label must also be collected.
5. The following things must be collected from the teacher:
   a. University
   b. Subject: Associated with a subject ID
   c. Chapter name: Associated with chapter ID for a given subject
   d. Question: Associated with a question ID
   e. Answer key: Mapped the question ID
      i. It should be a complete answer with proper sentence construction. Small keywords are not allowed while collecting the answer
      ii. All the points that are expected to be in the answer must be well written in each sentence.
   f. Keywords: Mapped with a question ID
      i. All the keywords that are expected to be present in the answer must be present in the database as comma separated values.
      ii. It can be shown in the User Interface using the add button and can be passed as comma separated while calling the backend API.
   g. All the questions and answer keys can be added one by one.
6. The following things must be collected from the students:
   a. The Question ID: qid
   b. The answer
   c. The Relevancy Score allotted by the teacher

Note: The mark is not the target variable. Mark varies from university to department etc. The relevancy score is what we are looking at.

7. Avoid unwanted Bias that are induced by the teacher:
   a. For the same answer two teachers must not give different relevancy score
   b. It should be captured from a single teacher with very high experience
   c. The corresponding subject teacher must sit with our analytical team to tell us all the aspect that they look in answer w.r.t to the key and score them.
   d. This will help in feature engineering. The teacher should stick with the procedure of relevancy scoring strictly so that bias is not induced in the dataset. If not, the ML algorithms will underperform due to ambiguous situations.

e. Hence there must be only one teacher who are Subject matter expertise who must annotate the relevancy score for each answer. answer w.r.t to the key and score them.

f. This will help in feature engineering. The teacher should stick with the procedure of relevancy scoring strictly so that bias is not induced in the dataset. If not, the ML algorithms will underperform due to ambiguous situations.

How do we capture answers in colleges/Universities?

If we are trying to do the same stuff in college with the help of teachers, we can automate the examination process of the CORE subjects where the number of retentive answers is high. Moreover, the data collection process will happen for all the units within a semester itself.

Remove the roll number, name, department as done in the school and hire someone data entry people to Type it and feed it into the database. **Funding is needed in this process. Director's approval is needed.**

DATA ANALYSIS, MODEL BUILDING AND FEATURE ENGINEERING

The inspiration for solving this problem comes from the Quora Question Similarity challenge that was hosted by Kaggle. The objective was to build and ML algorithms that can efficiently get similar questions to the given question i.e. Asked by the user.

In our context we are trying to find the answer key and answer similarity.

The most important thing in solving this problem is feature engineering. The features must be :

- subject specific
- Must capture all ends of the teacher's intuition
- Must capture, sentence construction, word similarity , sentence similarity, keyword mapping, length of answers all of these that must be fed into the ML algorithm to classify the relevancy score.
- Even though the features will be very generic but still the distribution of features with respect to the relevancy score will change from Subject to Subject. Hence custom models must be build on each subject.
- One single model that can do the scoring for all the subjects is not possible.

The reference to the challenge is given below : Please visit the kernels to get the best features that have improved the accuracy of the algorithm.

Competition details:

https://www.kaggle.com/c/quora-question-pairs/data

Top kernels :

https://www.kaggle.com/c/quora-question-pairs/kernels

Project Takeaways: Concepts that will be applied For Data Science Team

- Sampling and bias reduction
- Upsampling and Downsampling
- NLP feature engineering, word2vec embedding and text preprocessing
- Machine learning Quality Analysis
- Model building and finalization: All ML algorithms
- AB testing and Residual Analysis
- Production sing Models in web application using Docker and containers

----

Web application overall features:

- General user access controls for teacher and student
- Teacher dashboard for adding the questions and keys chapter wise for MCQ , as well as retentive questions
- Create test for the students for given class: should be there for a given time. Expire after a given day.
- Students login, access the test, give the test, MCQ and type the retentive questions in text boxes and submit
- The python service will invoke fetch the answers batch by batch and score the answers with respect to the keys.
- Student can login to see the results in an electronic report card manner.
- This is the overall view of the app to build a Proof of concept or MVP.

Web Development Domain Takeaways:

1. Angular Front End building
2. Node BackEnd building
3. Containerization using Docker
4. Working in cloud environment. Azure or AWS
5. Building a logging system using Elasticsearch or MongoDB
6. Quality analysis and testing

Team needed :

2 backend : Nodejs , Docker ,Node ORM for MongoDB and MySQL

1 Frontend : Angular or React

<div align="center">**Exam Automation : Key points and difficulties**</div>

1. **Question framing policies and examples :**

The objective is to keep the questions whose answers have very less subjectivity. We will be forcing certain aspects of the memory of a student where it is important for them to memorise certain answers and process. The question which need their analytical reasoning or aptitude we can put them in MCQ questions. The system should equally tap in the memory and thinking ability of a student. That is the system we are going to develop and that's where the automation can kick in perfectly.

Examples:

Definitions:

- What is absorption?
- What is Apical meristem and what does it do?
- What is differentiation in tissues?
- What is radiation?

Process explanations or functionality explanations:

Dividing the whole problem into sub questions where the objectivity is very high and try to find the similarity for each sub section and aggregate the whole scoring process.

Explain the following:

- Muscular tissue:
- Connective tissue:
- Nervous tissue:

Instead of asking what the types of thermal energy flow are we can convert into: explain the following process:

- Conduction
- Convection
- Radiation

Edge cases: Extreme cases.

Answer key:

Plants inhale carbon dioxide and exhale oxygen during the day, during the night plants inhale oxygen and exhale carbon dioxide.

Answers:

Plants exhale carbon dioxide and inhale oxygen during the day and during the night it is vice-versa. 0 relevancy score

Correct answer: 5 relevancy score but the features will distribution will be different.

Plants inhale carbon dioxide and exhale oxygen during the night, during the day plants inhale carbon dioxide and exhale oxygen.

Plants inhale carbon dioxide and exhale oxygen during the night and in the day its vice versa.

Correct answer : 5

How to account for this variance and sentence construction ?

- active / passive voice , the subject and the object interchanges : challenge
- correct answer with same sequence : 5
- correct answer with different sequence : 5

Need to discuss with professors if we can transfer such type of questions into MCQ format

Difficulties in Data collection:

- Conversion of handwritten answers to digital text.
- Making sure the entered data is free from spelling mistakes and other errors.
- Framing a standard policy to annotate key words for every answer.
- Access to answer scripts.
- Cooperation of the university to allow us to have a demo run to collect data for custom set questions with 1st Years.
- Teacher Support on all levels : Setting of Questions and assigning relevancy scores.

Relevancy Score annotating policies:

Constraints in the model:

- **No formulas**
- **No equations**
- No spelling mistakes i.e. if the keywords spelling do not match in answer key and answer even though its similar then it's a problem. A possible workaround can be regular expressions etc.
- Difference of voices of Answer key and Answer(active/passive)

The Relevancy annotation can vary from subject to subject:

1. First we need to identify what are the things we want to enforce to the students.
2. Look whether they are touching keywords
3. Look whether they are touching keypoints and covering the right facts
4. Look whether they are constructing the sentences right i.e. Optional
5. Check whether all the points are covered

Note: *We need to remember that the relevancy scores which are given by the teacher will be used to predict. In any general machine learning scenario we will have a predictor which either occurs naturally*

or is a part of business scenario. We engineer features to explain the variance and predict the as accurate as possible.

In our case we are the ones who are to annotate the relevancy score. Based upon the factors which we want to enforce , if the relevancy score is given we can engineer features accordingly and explain the variance of the relevancy scores (0-5)

Hence the relevancy score annotating policy varies from subject to subject and it should be agreed by both the parties.

Technical Requirements:

- Stable Internet Access (Dedicated to the team)
- Dedicated storage space
- A room available 24/7 for the team to work whenever possible

Non-technical Requirements:

- Cooperation from respective subject teachers to help us in setting questions (only first years)
- On Duty permissions for the small team working this semester for data collection and operations.