Bonus Assignment

"Finding Potential Mining Sites for Naturally Occurring Metals Using Explainable NLP"

Group Member 1: **Gitansh Raj Satija** Group Member 2: **Pritish Wadhwa**

Ans 1.

- **Explainability Components**
  - Explainability components can be implemented at various stages of the pipeline to better understand the actualities of the model.
  - This presents us with some sort of pseudo-explainability, which more often than not is sufficient to get a gist of the model.
  - eg. in decision trees, we can get to know at each node as to how the split is going to be.
  - Continuing with Mr. Dash's project as an example, he added multiple explainability components to various stages which gave some sort of confidence to the predictions by highlighting the title keyword phrases and the abstract keyword phrases.
- **Feature Selection and Extraction**
  - For tasks like Mr. Ayushman's Project, or other NLP-based projects, keywords and key phrases play an important role in the explainability component of the model. Similarly, apart from this, we can display the importance that the model is giving to all the different features it is being trained on.
  - By this, we get to know which of the features play a more important role in predicting output for a particular sample and how each piece of information is influencing the outcome.
- **Human in the Loop**
  - Given enough resources, we can also have domain experts who can rank the important features, phrases, keywords that play a role in the output of a sample will provide us with data which we can feed our model using which it can learn about which aspects play a higher role in getting to the output.
  - Similarly, the explainability that the models give, can be verified by domain experts, and feeding this information back to the model can help improve its performance.
- **Periodic Result Sharing**
  - It becomes essential to keep the stakeholders involved in the process. They usually have little knowledge of the algorithms behind the model, thus explaining to them what is going on behind the black box of a model becomes a real challenge.
  - Even though overcoming this is a daunting task, usually it can be somewhat achieved by keeping them in the loop by giving them weekly updates about the results so that they might also get a gist of what is going on. This is important as after all they are the ones funding/backing your project :P

Ans 2.

- **Precision**, in short, is the measure of **quality**. It is the number of relevant instances (true positives) amongst all the retrieved instances (all the samples classified as positive instances(true positives + false positives)).
- An ML model (like Mr. Ayushman's project) can have a **confidence score** for every prediction that it makes. Higher confidence implies that the model is more sure about the prediction that it has made than a prediction. We can appropriately set a **confidence threshold** with which we want our model to predict something as positive. Thus by varying a confidence threshold (which can be done using an **ROC curve**) we can control our precision.
- By **Controlling precision** we can control the quality of predictions we want from our model. High precision (higher confidence threshold) implies that we have a high true positive rate i.e. whatever our model is predicting as positive it is very likely to be positive. This can be especially useful in tasks like this project which want very low false positives. However, there can be a possibility that for a project the expectation is to capture as much positive (i.e. classified as 1) data as possible without worrying about some false predictions that can be there. Thus for this, we'll want a low precision (or lower confidence threshold). By doing this we are focussing more on **quantity** than quality.

Ans 3. In any ML project, the involvement of humans can be beneficial in all the phases from dataset collection to testing the model predictions.

- **Dataset Collection**
    - A lot of ML projects like Speaker Detection, Human Activity Recognition among many others require raw data which humans can provide.
    - Many tasks require annotations for which we require a lot of human support (as well as domain experts) who can help with these annotations which are critical for human feedback. Some tasks apart from the NRCan project that requires annotation can be hate speech detection among others.
- **During Testing Phase**
    - Every real-life ML project initially uses a limited training and testing set and is never able to give 100% correct predictions. However, when the model is deployed for use, we want its performance to be as high as possible. Thus, verifying the predictions made by the model can indeed help in improving its performance by feeding this data back to the model. This can be done by either human domain experts (for tasks like Mr. Ayushman's project) or even by non-experts depending on the task.

Humans in the Loop is a very critical aspect. On building large-scale ML Models like Handwriting to Text Conversion, Chatbots, or other tasks it is important to have as many humans involved in the data collection as possible to provide scalability to the model and make it free of any biases. Similarly, at every phase active human involvement can help **improve the performance of the model**.

Ans 4. The project discussed had a lot of strengths associated with it. Some of them are:

- **Adaptable Model**
    - He discussed that the system his team created could easily be changed for other metals also, thus making it highly adaptable.
- **Human in Loop**
    - Active human involvement during different phases of the project can help improve the model performance, In the case of NRCan, the involvement of domain experts in verifying the predictions made by the model is a critical aspect as it helps the model to know if it is performing correctly and in case of any wrong predictions, it can improve itself.
- **Deployment Strategy**
    - He used Canary deployment as the deployment strategy for his APIs. This was one of the things that impressed and awed us the most as it makes sure that there are no issues to the end-user at any time. The gradual shift in traffic makes certain that the system is perfect, thus preventing sleepless nights for both the developer and the manager.
- **Credible Data**
    - As Jainendra Sir has said a lot many times, "your model is as good as the data you have". Mr. Ayushman also got credible data for his project. Even though he got only 1000 files (he could have gotten 1000 more had he required), those were written by domain experts. There were some false positives in the tags, but more or less the data he had was credible enough to make sound predictions.

Ans 5. The project he presented was amazing. It was very difficult to find any mistakes or shortcomings in it, still, some points that bugged us later were:

- **Availability of Geolocation**
    - He mentioned in the talk that the geotags available to them had a false positive rate of as high as 70%. We believe that these geotags can be very useful as they can provide some sort of credible information about the actual site. He could have used technologies like satellite imaging to get to know the land a bit better, eg. a place that has rich coal mines would have a certain kind of vegetation and soil around it. These features would definitely make his results better and might even make the model more explainable.
- **Improving OCR for full-text Extraction**
    - Advanced image processing techniques including image distortion can be used to improve the performance of OCR's on old documents which will help in extracting the full text of the documents as well. This text will most likely help build a better model and also help improve the explainability component of the model.