PLP PROJECT INDIVIDUAL REPORT & PEER REVIEW

Project: Community Help

OCTOBER 15, 2023

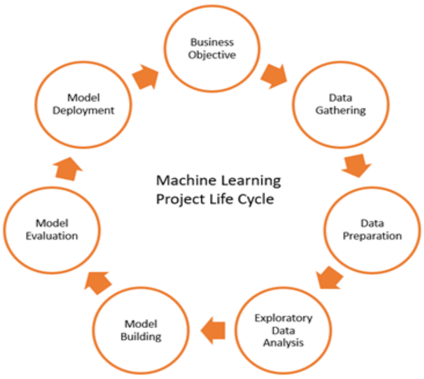
NATIONAL UNIVERSITY OF SINGAPORE

Student Name	Thota Siva Krishna Vara Prasad
Email	Sivakrishnathota5@gmail.com
Project Title	Community Hep **NLP Components:** 1. News Classification. 2. Sentiment Analysis. 3. Machine Translation Hindi to English. 4. FAQ Chatbot.

My contributions to the project:

1. Dataset collection for news classification, Sentiment Analysis, Machine Translation.
2. Prepared Data for FAQ Chatbot.
3. Data Cleaning pipeline.
4. Convert Text to features using Bag of words, TFIDF.
5. Train Deep Learning and Machine Learning models for news classification.
6. Train Deep Learning and Machine Learning models for sentiment Analysis.
7. Train Deep Learning and Machine Learning models for Machine Translation & FAQ Chatbot.
8. Model Evaluations & Comparisons.
9. Predictions using real time data.
10. Prepared project reports & video presentations.
11. Develop Community help website using python & Django framework.
12. Designed conversational UI for FAQ Chatbot using python and flask framework.
13. Designed conversational UI for Translation chatbot using python and flask framework.

While designing and developing projects, I followed all steps in Machine Learning project life cycle.



Self-reflections:

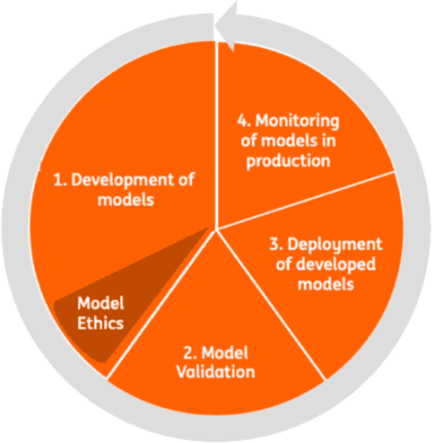
What is NLP?

Natural Language Processing (NLP) is a part of computer science and artificial intelligence (AI) which gives the machines the ability to read, understand and derive the meaning from human language. This ability of machines facilitates many services which we use in our daily life maybe without noticing. When you type half of the word while chatting, nowadays all smart phones can complete your words before you finish it. There is an automatic grammar corrector in most email providers as well. All these tools have an NLP algorithm behind the scenes.

If you have a business use-case where you need to build an NLP model, how would you start and end the lifecycle of the model?

Let's address first a business use case in which we use News classification, sentiment analysis. The challenge to deal within this use case is evaluating unstructured and structured text tweets in the most efficient way. Having an automatized evaluation system with NLP models that assigns topic and sentiment of the feedback is considered as the solution to this challenge.

To achieve this aim, we have built a topic and sentiment model separately by using NLP. I would like to give you a general sense about all stages of the full life cycle of models and main takeaways from what we experienced during our journey. We consider 4 main stages of the full life cycle to build and maintain topic and sentiment models: development, validation, deployment, and monitoring of the models.


4. Monitoring of models in production
1. Development of models
3. Deployment of developed models
Model Ethics
2. Model Validation

1. Development of the models

Development stage is the first focus and probably the most time-consuming stage. It starts with the designing and developing the modelling steps which include data, methodology and performance metrics by considering the limitations.

To give a more solid explanation, let's focus on our sentiment model and how we develop it.

Data:

We have structured text data with some irrelevant or sensitive info (e.g., emails, corporate keys) We cleaned the text data first.

Methodology:

There are supervised, semi-supervised and unsupervised approaches you can use to predict the sentiment of text data. We started with a supervised method since we have annotated data.

Solution:

Starting with the simple approach and switching to a more complex and time consuming.

Having the correct sentiment for a small group didn't work very well since there wasn't enough data to get the pattern by the model.

Performance metric:

To compare different models and ensure that the model is working sufficiently, you must define a solid metric to measure. There are multiple options (e.g., precision, recall) and you should choose based on your intended usage of model output. In our case, we used the f1-score which is the harmonic mean of precision and recall considering both false positive and negative cases.

1.1 Model ethics

In addition to technical details, there is also the non-technical aspect of the development phase which is model ethics. Ethical and moral issues are very important to investigate to be sure that the model doesn't have any bias on specific group(s) (e.g., gender, language, or country etc.). We should address the following questions during this investigation:

- Does the model make more mistakes for a specific language?
- Does the model have the ability to detect gender or nationality of the respondents and use this information while making sentiment prediction?

Here are some suggestions to address these questions:

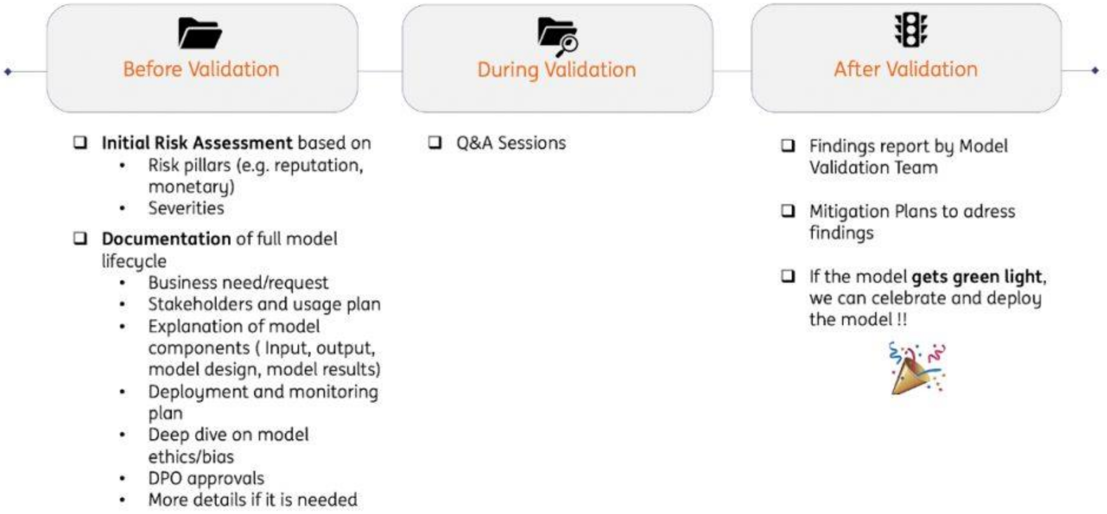
- Performing error Analysis per specific group (e.g., language) to see if the model has significantly lower performance for any group.
- Building another model to predict specific group from the feedback (e.g., gender) and checking if the performance is good, meaning that the model can derive the gender by only looking at feedbacks

Key takeaways:

- Start simple as long as it covers the need.
- Iterate the development by improving something in every step.
- Keep in mind the limitation of the use case and the design of the steps of development accordingly!
- Deep dive model results to investigate technical and non-technical aspects.

2. Model validation

Since Community help as a Social Network is in a highly reputation industry, we must validate developed models before deploying them in production. So far, Community Help has established a very well-structured model validation framework which is summarized below.



□ **Initial Risk Assessment** based on
- Risk pillars (e.g. reputation, monetary)
- Severities

□ **Documentation** of full model lifecycle
- Business need/request
- Stakeholders and usage plan
- Explanation of model components (Input, output, model design, model results)
- Deployment and monitoring plan
- Deep dive on model ethics/bias
- DPO approvals
- More details if it is needed

□ Q&A Sessions

□ Findings report by Model Validation Team

□ Mitigation Plans to adress findings

□ If the model **gets green light**, we can celebrate and deploy the model !! 🎉

Key takeaways:

- Be aware of validation requirements while designing the model.
- Document every detail while developing the model (e.g., training and test set, detailed results, and explanation)
- Plan the deployment and monitoring stages before starting the validation.

3. Deployment of developed models

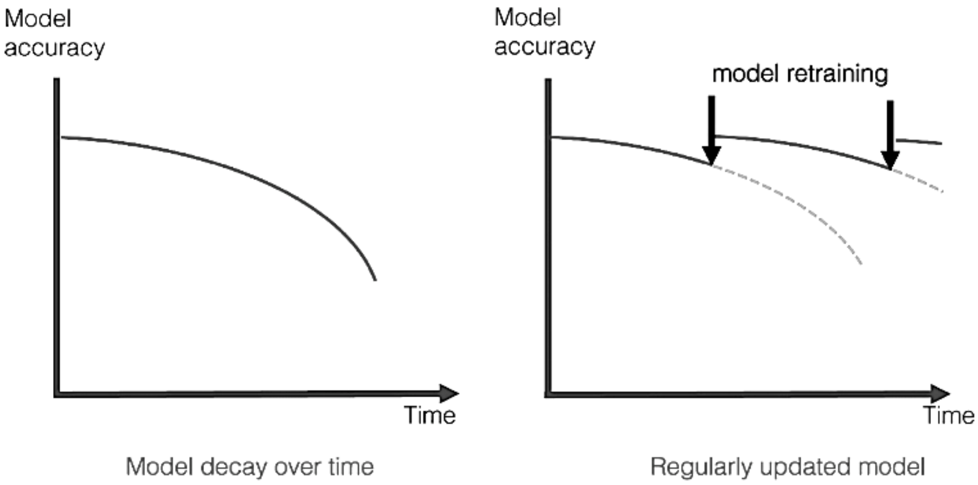
Once you have finished the development and are sure that it is a valid model, you save the trained model in a re-callable file format and deploy this model in the production to get predictions on new data. You should follow the same data preparation steps to help the model to see feedback in the same standards and call the saved trained model to make a prediction for new data during the deployment. If you conduct a new survey (meaning new Tweets data) in a consistent frequency, you can automatize this process.

Key takeaways:

- Apply the same preparation steps in the development stage on new data before getting predictions.
- Automate the deployment based on frequency of Tweets.

4. Monitoring of models in production

Models tend to be obsolete and suffer performance drop over time by their nature. This is called model decay. Once it has started, the retraining of the model must be done to maintain the performance of the model at a certain level. Monitoring is essential to detect this retraining need on time to avoid model decay.



Depending on the use case, you must plan the monitoring stage and once the model has been deployed in the production, you should activate a monitoring system as well. We can categorize use cases into:

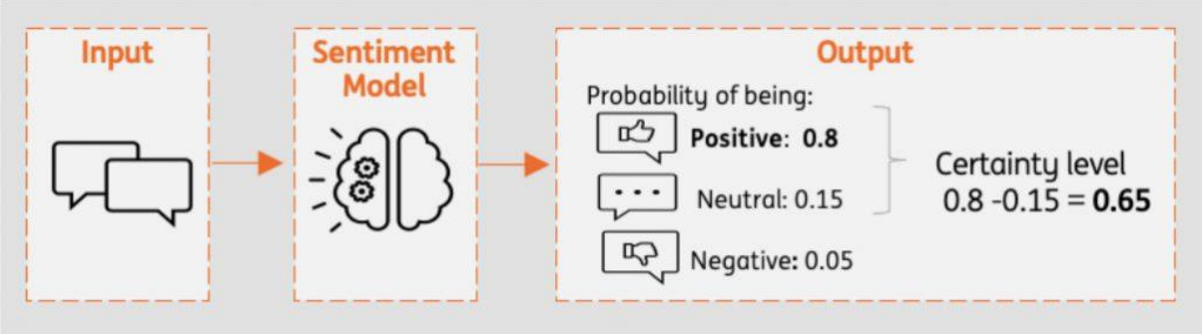
Case 1: Available correct labels after making prediction.

Monitoring Method: Check the performance metric between predictions and correct labels over time.

Case 2: No luxury to know labels without manual annotation.

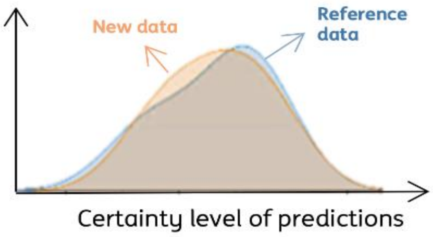
- Monitoring Method: Novel approach called drift detection methods on defined variables.

Our business use case is placed in case 1 since we have the luxury to know labels without manual evaluation every time when we use the model on new Tweets data. That's why we use drift detection methods.



Drift detection methods track the distributional shift in a defined variable for two different datasets. For us, these two datasets are training data as reference data and new data in production. We define the variable which we would like to track for a shift as a certainty level of predictions. Predictions are made based on probabilities of being positive, negative, and neutral in the sentiment model. The sentiment with the highest probability is chosen as predicted sentiment. We calculate the certainty level as the probability difference between first two class probabilities.

If there is a significant change towards left, it means that there is a shift and retraining need!



After establishing the drift detection method with these details, we perform an evaluation experiment on the monitoring system. We apply the method on new data and check if the method concludes with a drift. In parallel, we annotated manually a small amount of feedback from new data and checked if there is significant change on performance metric. According to the result, we ensured that the established monitoring system is working.

Key takeaways:

- Establish a monitoring system depending on the use case.
- Have a proper test on the monitoring system designed before using it!

Conclusion

explained the key points of each stage in the full model life cycle based on what we experienced in our journey.

The most important conclusion especially for the experts who are at the beginning of this journey is that building NLP models (or any machine learning models) doesn't mean only training a model which gives predictions with the best performance. If you would like to maintain the impact of the model in the long term on business use cases, you must consider the full life cycle of the model.

Peer Review

Team Member Name	Contribution Factor (an integer from 1 to 5; 1 = poor, 5 = excellent)	Comments
Thota Siva Krishna Vara Prasad	5	Involved in all Level of component design, Development, and documents preparations