

Essay Writing Feedback

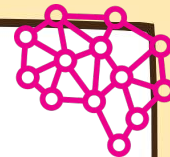


Shringa Bais
May 2023

(From Kaggle Competition - Predicting Effective Arguments)

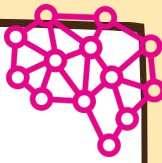


Agenda



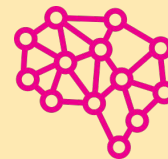
- Problem Statement
- Current challenges & its impact
- Solution & Assumptions
- Success Metrics
- ML Model Pipeline
- Demo
- Application Architecture
- Responsible AI
- Achievements & Future Work

Motivation



Effective writing is critical for success in college and future careers, but few students graduate high school as proficient writers. According to the National Assessment of Educational Progress (NAEP), less than a third of high school seniors are proficient writers. That is especially true within marginalized communities of low-income, Black, and Hispanic students – less than 15% score proficient (NAEP).

Challenges & Impact



Challenge

Time consuming for teachers to provide necessary feedback for a student growth.

Impact

Students does not receive effective feedback to become a confident writer



Challenge

Present solutions are expensive and proprietary.

Impact

Less accessible for educators and students who needs it mostly



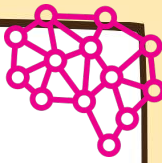
Challenge

Existing solutions are immature with limited accuracy. Helps to resolve only grammatical errors.

Impact

Students ability to improve writing skills and potentially hinders technology's adoption

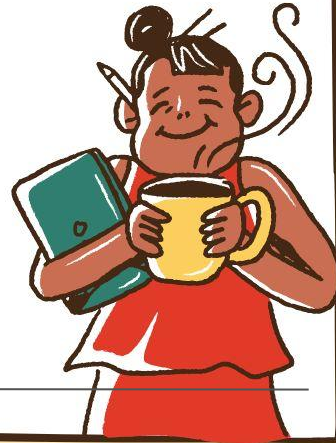
Proposed Solution



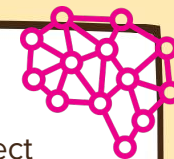
Develop an Assisted Writing Feedback Tool (AWFT) for educators to give effective, personalized and timely feedback on student writing.

Applications:

- Streamlining the Grading Process
- Empowering Teachers
- Maximizing Student Potential

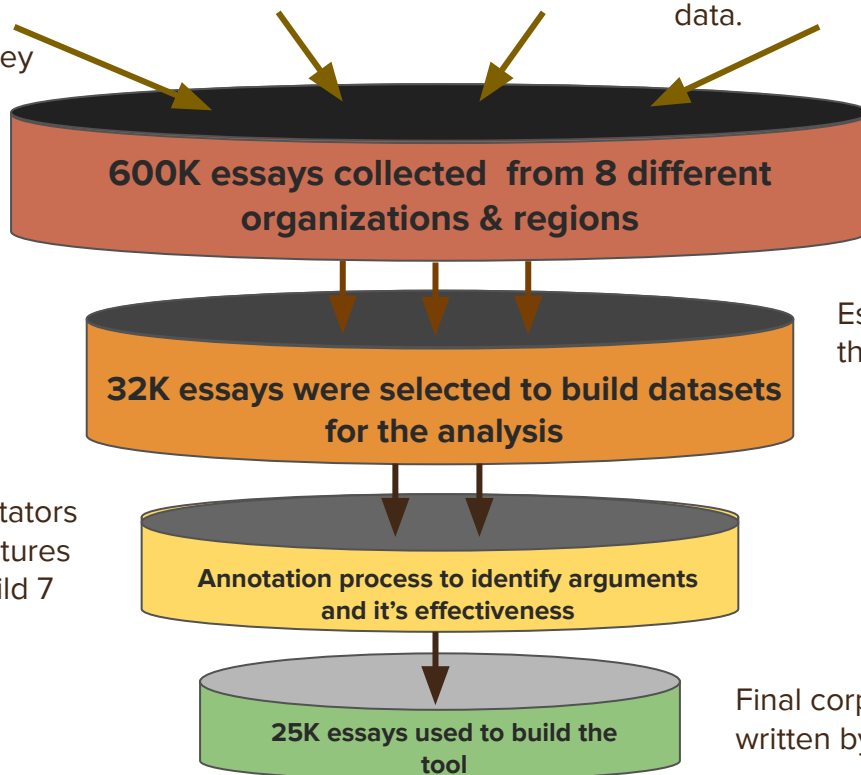


Assumptions based on Surveys



Various educators and academic organizations took part in the survey and provided feedback.

Suggestions and thorough discussions helped identify the right group and collect data.

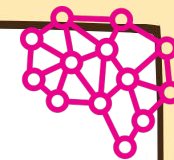


Essays were selected based on the teacher advisory boards

Educators along with annotators build an argumentative features to annotate essays and build 7 unique elements to data.

Final corpus represents the essay written by students in 6-12 grades.

Expectations



Many people believe that the Electoral College should be abolished, while others believe that the Electoral College should stay as it is.

However, what most people who want to keep the electoral college, do not know is that when you vote for a presidential candidate you are actually voting for a slate of electors, who in turn elect the president. Which means that the people do not get a direct vote towards the president. Therefore, it can cause disinterest in people who are eligible to vote. That is why I argue in favor of changing to election by popular vote for the president of the United States.



Many people believe that the Electoral College should be abolished, while others believe that the Electoral College should stay as it is.

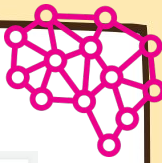
However, what most people who want to keep the electoral college, do not know is that when

you vote for a presidential candidate you are actually voting for a slate of electors, who in turn elect the president. Which means that the people do not get a direct vote towards the president.

Therefore, it can cause disinterest in people who are eligible to vote. That is why I argue in favor of changing to election by popular vote for the president of the United States.

**Lead
Argument
- Effective**

Success Metrics + Monitoring



- Number of unique essays.
- Increase in tool adoption i.e. number of predictions over time.
- Number of teachers on the platform
- Low processing time $\sim 10\text{ms}$

CPU Utilization: Average

Percent

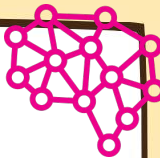


predict_event

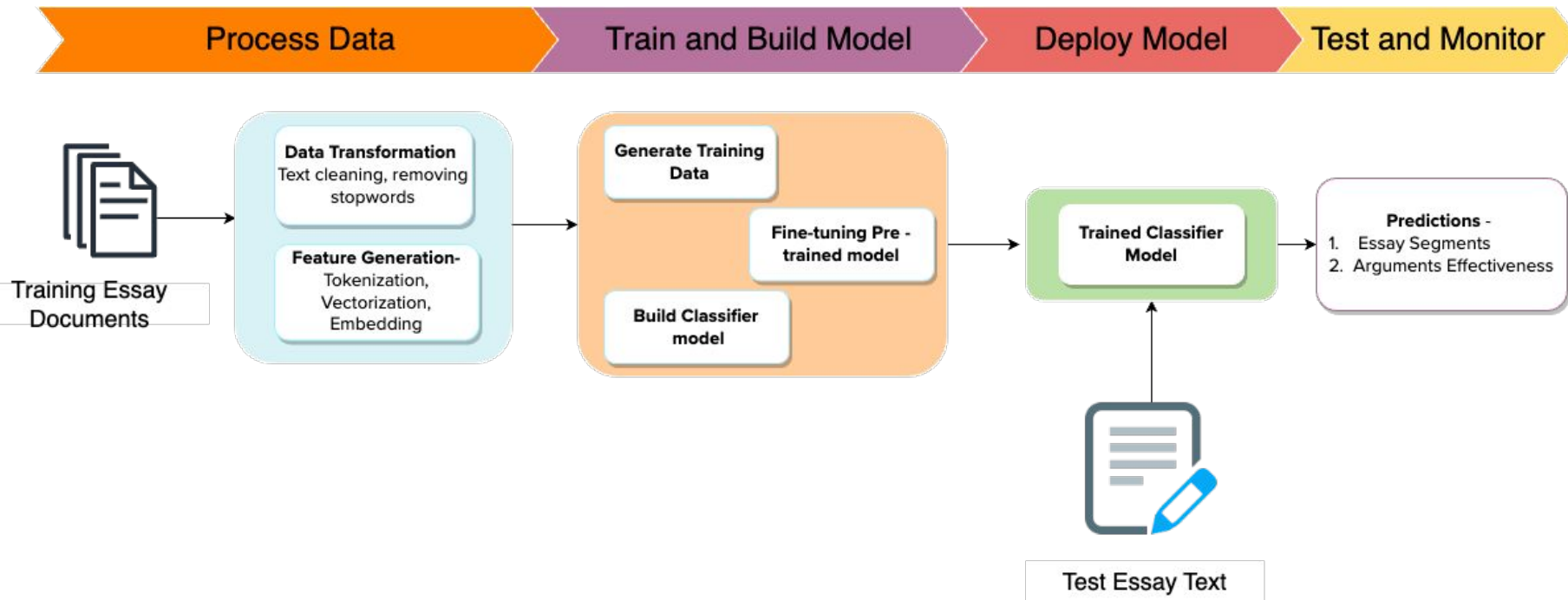
7

predict_event

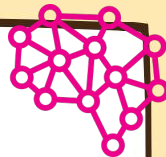




ML Model Pipeline



Demo



Enter Essay Text

Hi, i'm Isaac, i'm going to be writing about how this face on Mars is a natural landform or if there is life on Mars that made it. The story is about how NASA took a picture of Mars and a face was seen on the planet. NASA doesn't know if the landform was created by life on Mars, or if it is just a natural landform. On my perspective, I think that the face is a natural landform because I dont think that there is any life on Mars. In these next few paragraphs, I'll be talking about how I think that is is a natural landform

I think that the face is a natural landform because there is no life on Mars that we have discovered yet. If life was on Mars, we would know by now. The reason why I think it is a natural landform because, nobody live on Mars in order to create the figure. It says in paragraph 9, "It's not easy to target Cydonia," in which he is saying that its not easy to know if it is a natural landform at this point. In all that they're saying, its probably a natural landform.

People thought that the face was formed by aliens because they thought that there was life on Mars. though some say that life on Mars does exist, I think that there is no life on Mars.

It says in paragraph 7, on April 5, 1998, Mars Global Surveyor flew over Cydonia for the first time. Michael Malin took a picture of Mars with his Orbiter Camera, that the face was a natural landform. Everyone who thought it was made by aliens even though it wasn't, was not satisfied. I think they were not satisfied because they have thought since 1976 that it was really formed by aliens.

Though people were not satified about how the landform was a natural landform in all we new that aliens did not

Clear

Submit

output

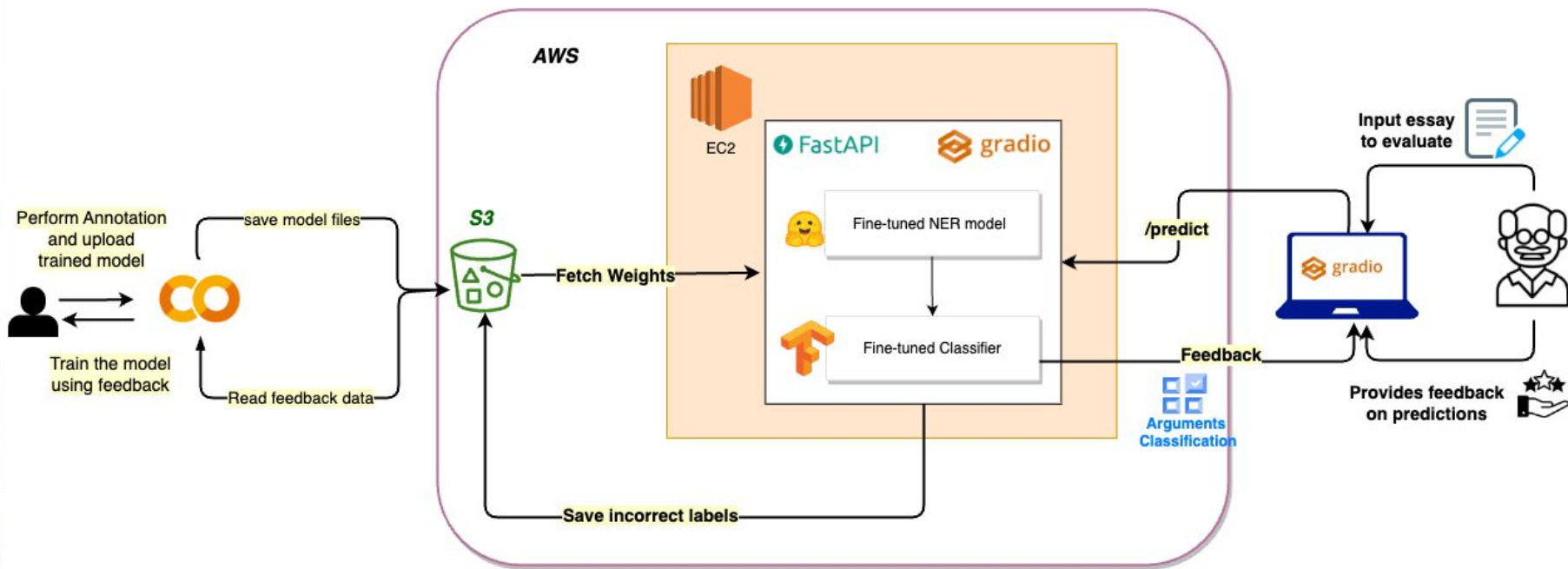
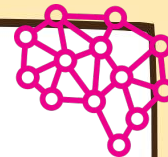
hi i m isaac i m going to be writing about how this face on mars is a natural landform or if there is life on mars that made it the story is about how nasa took a picture of mars and a face was seen on the planet nasa doesn t know if the landform was created by life on mars or if it is just a natural landform **LEAD:: ADEQUATE** on my perspective i think that the face is a natural landform because i dont think that there is any life on mars in these next few paragraphs i ll be talking about how i think that is is a natural landform **POSITION:: ADEQUATE** i think that the face is a natural landform because there is no life on mars that we have discovered yet **CLAIM:: ADEQUATE** if life was on mars we would know by now the reason why i think it is a natural landform because nobody live on mars in order to create the figure it says in paragraph it not easy to target cydonia in which he is saying that its not easy to know if it is a natural landform at this point in all that they are saying its probably a natural landform **EVIDENCE:: ADEQUATE** people thought that the face was formed by aliens because they thought that there was life on mars **COUNTERCLAIM:: ADEQUATE** though some say that life on mars does exist i think that there is no life on mars **REBUTTAL:: ADEQUATE** it says in paragraph on april mars global surveyor flew over cydonia for the first time michael malin took a picture of mars with his orbiter camera that the face was a natural landform **EVIDENCE:: ADEQUATE** everyone who thought it was made by aliens even though it wasn t was not satisfied i think they were not satisfied because they have thought since that it was really formed by aliens **COUNTERCLAIM:: ADEQUATE** though people were not satified about how the landform was a natural landform in all we new that aliens did not form the face i would like to know how the landform was formed we know now that life on mars doesn t exist **CONCLUDING STATEMENT:: ADEQUATE**

Flag as Adequate

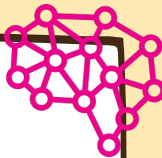
Flag as Effective

Flag as Ineffective

Application Architecture

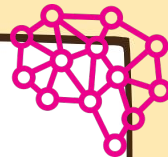


Achievements



- ✓ Real-time end-to-end essay feedback ML solution.
- ✓ Quantitative/Qualitative methods to classify essay arguments.
- ✓ Most language models suffer with large texts, average size of essay is >1K words, I used longformer architecture models to overcome this problem.
- ✓ Trained and Optimized embedding models for improved classification accuracy.
- ✓ Successfully deployed ML app on EC2 instance with an intuitive UI.
- ✓ Prediction results are self explainable and easy to understand.

Responsible AI



Privacy and Data Protection - Ensure handling of each student data privacy by handling data secure.

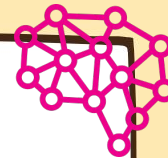


Collaborative and Diverse Development - Be more inclusive of diverse cultural, linguistic, or socioeconomic backgrounds.



Regulatory Compliance - Stay up-to-date with relevant laws and regulations regarding the used model and data linked with it.

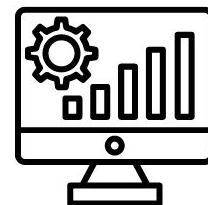
Future Work



Collect more data
to reduce imbalance data
effect



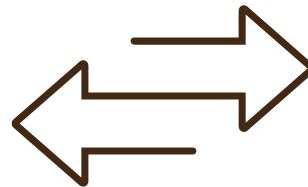
Continuous
improvement on the
model predictions



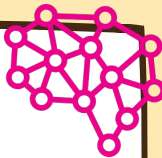
Monitoring system
to track individual
progress



Scalable solution to
handle multiple users



Two-way interface
for both students
and teachers



A Big Thank you to

Milica Cvetkovic

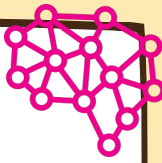
Chris Alexiuk

Anna Marek

Milan McGraw

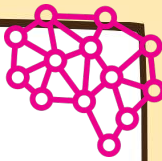
Sina Rafati

Questions??



Appendix

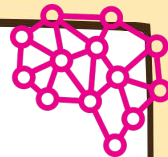
Data Stats



There are two different datasets for 2 different tasks:

- 1) **NER model** - To identify arguments in the data. The dataset contains 144K rows of labeled discourse data with required feature of discourse text length and its type.
 - a) Implemented baseline Longformer model through hugging face to create NER tokens and eventually train the model. It's an
- 2) **Classifier model** - To classify discourse arguments into 3 labels 'Adequate', 'Effective' , and 'Ineffective'. The dataset contains 36K rows of labeled argumentative discourse data with required feature of discourse type.
 - a) Started of with vectorizer approach to understand the baseline accuracy achievement and data learning.
 - b) Fine-tuned multiple pretrained transformer models such as BERT and DistilBERT.
 - c) Fine-tuned tensorflow hub model universal sentence encoder
 - d) Fine-tuned state-of-the-art python framework text embeddings model.

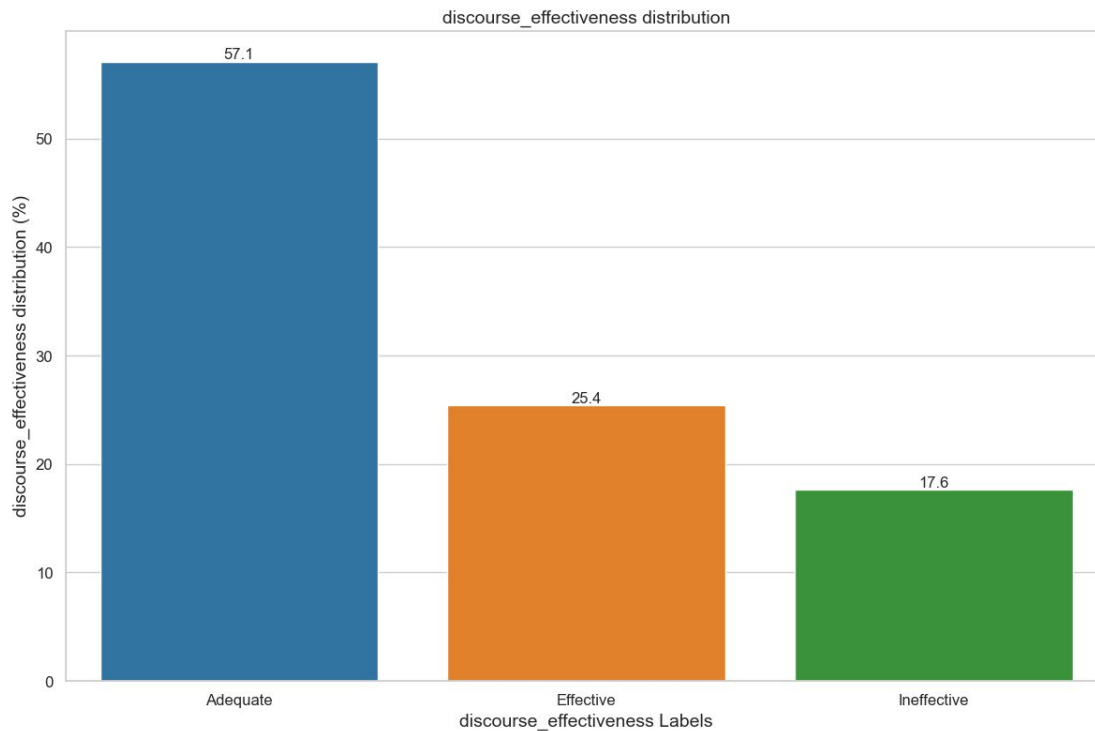
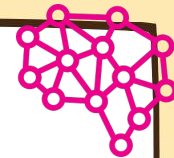
Discourse Types



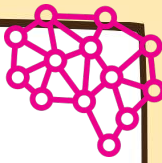
The dataset presented here contains argumentative essays written by U.S students in grades 6-12. These essays were annotated by expert raters for discourse elements commonly found in argumentative writing:

- **Lead** - an introduction that begins with a statistic, a quotation, a description, or some other device to grab the reader's attention and point toward the thesis
- **Position** - an opinion or conclusion on the main question
- **Claim** - a claim that supports the position
- **Counterclaim** - a claim that refutes another claim or gives an opposing reason to the position
- **Rebuttal** - a claim that refutes a counterclaim
- **Evidence** - ideas or examples that support claims, counterclaims, or rebuttals.
- **Concluding Statement** - a concluding statement that restates the claims

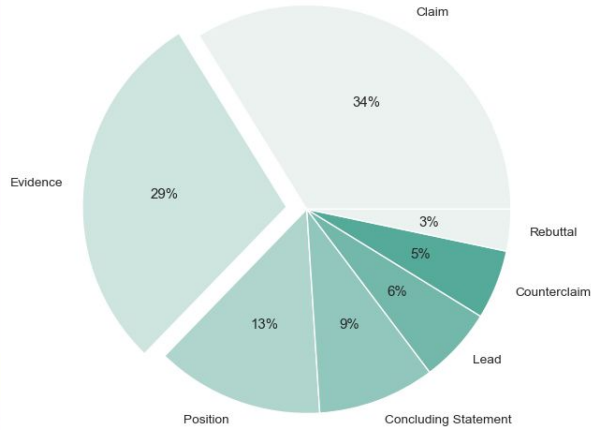
EDA - Target class overall distribution



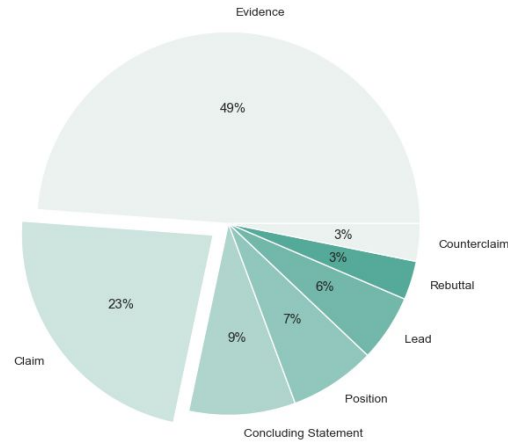
EDA - Target class by discourse type



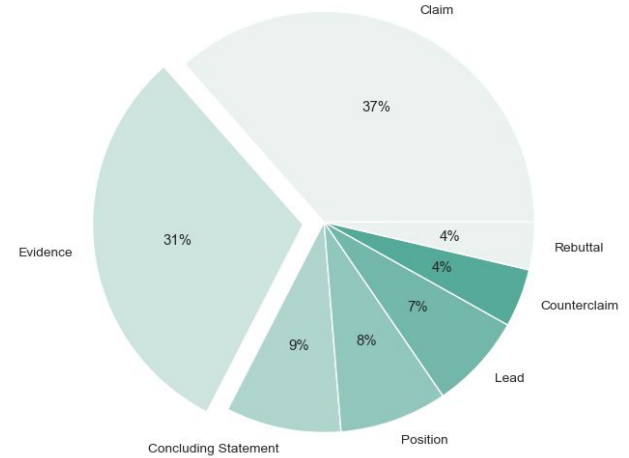
Adequate Argumentation Elements

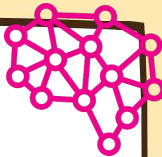


Ineffective Argumentation Elements



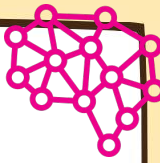
Effective Argumentation Elements



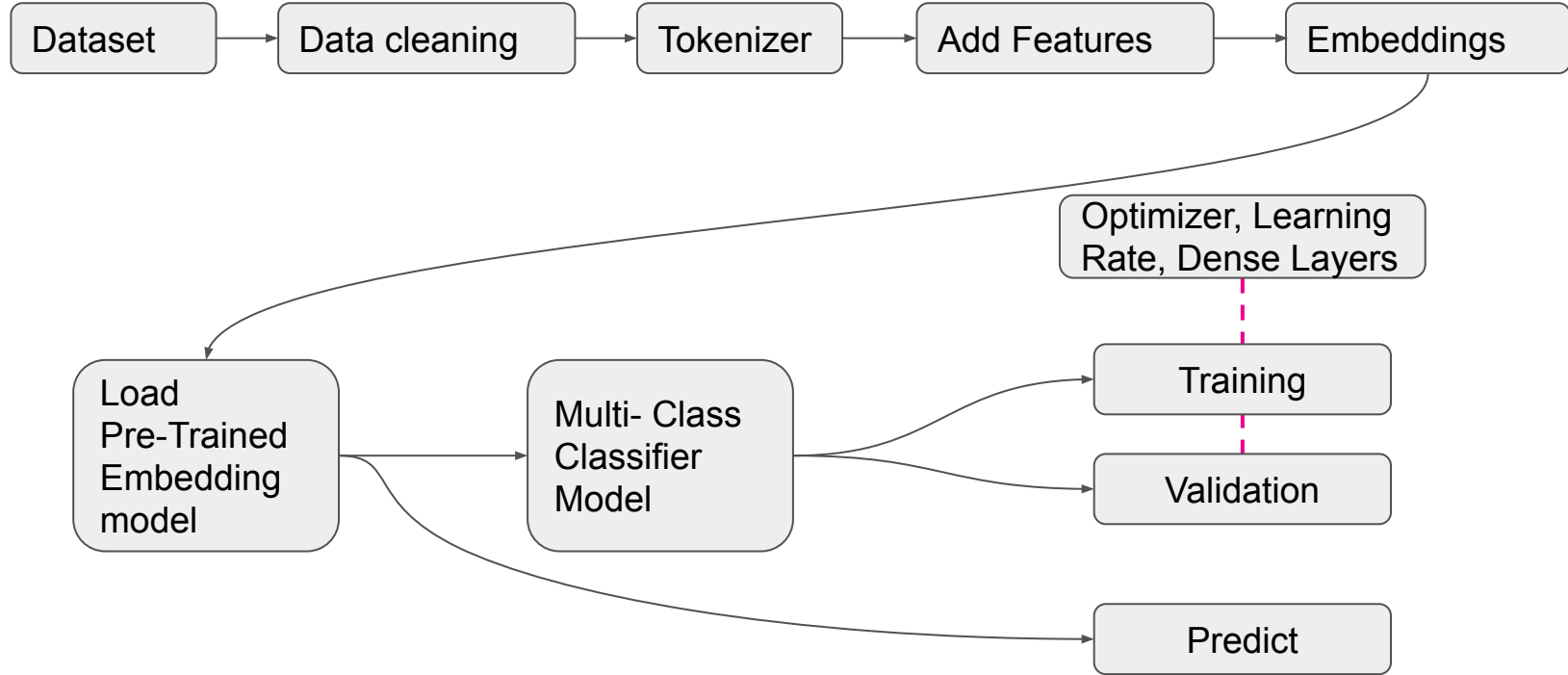


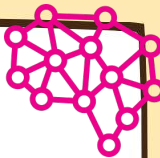
Feature Engineering

- Discourse Text
- Discourse Type + Tokenizer Separator + Discourse Text
- Discourse Type + Spacing + Discourse Text + Tokenizer Separator

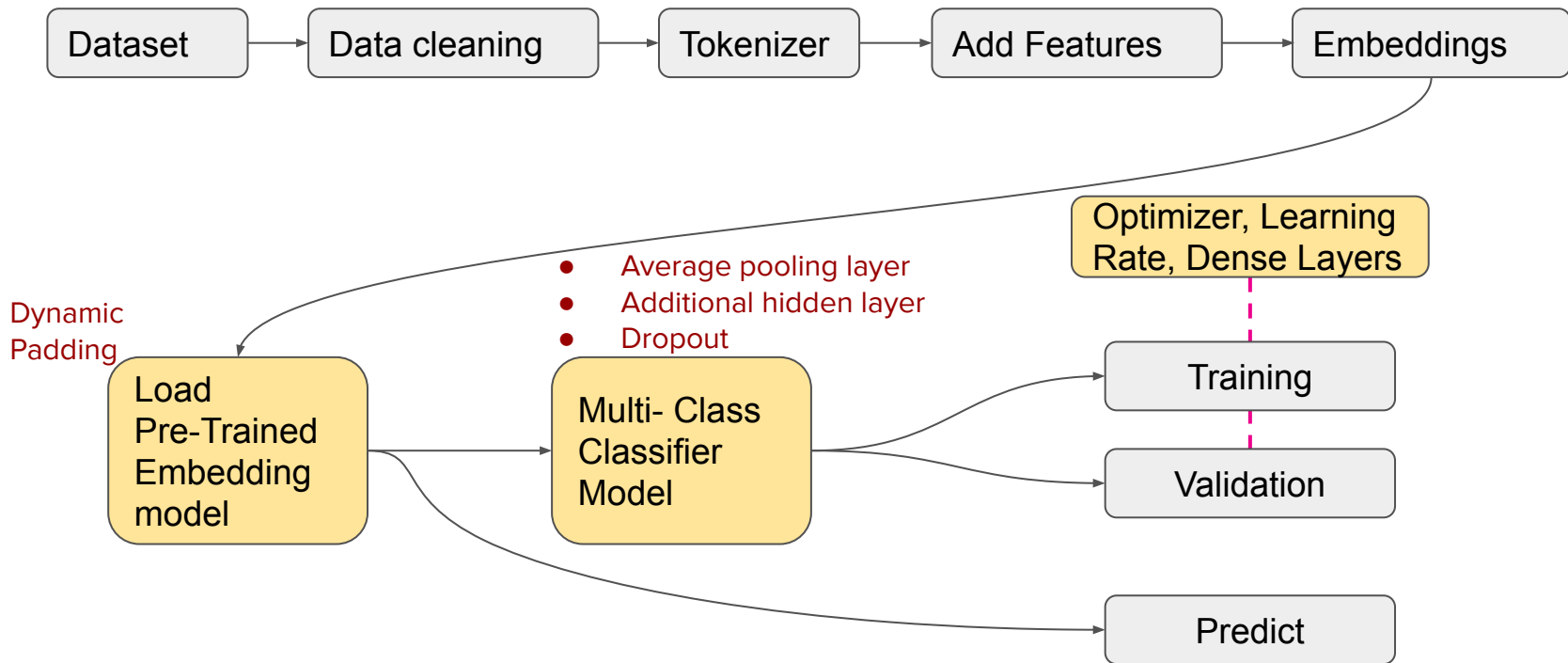


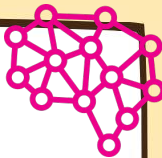
Deep Learning Model Flow





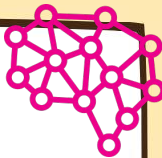
Hyperparameter Tuning





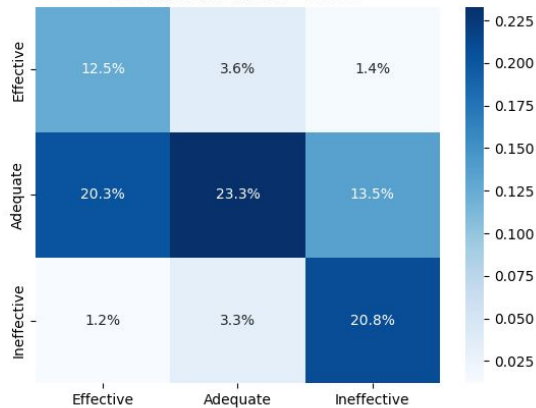
Evaluation Metric

- 1) F1 score (**Imbalanced dataset**)
- 2) AUC-ROC Curve (**Multi Class Classifier**)

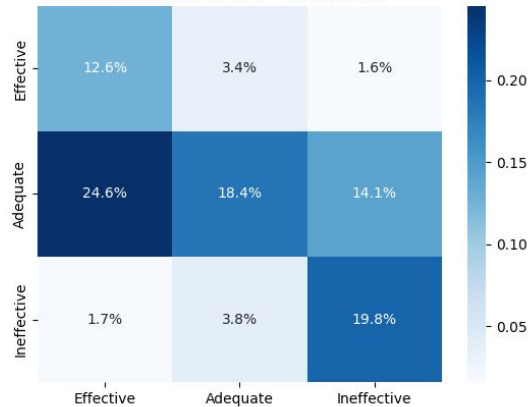


Results

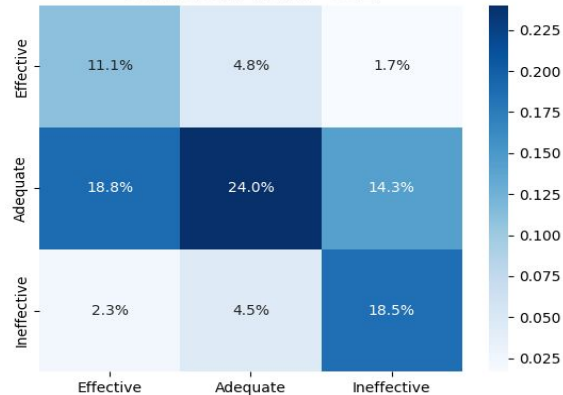
Confusion Matrix - BERT



Confusion Matrix - DistilBERT



Confusion Matrix - USE



Confusion Matrix - ST



Results

BERT performance is better than other models

F1 Score

Label	BERT	DistilBERT	Universal Sentence Encoder	Sentence Transformer
Effective	0.49	0.45	0.45	0.46
Adequate	0.53	0.44	0.53	0.69
Ineffective	0.68	0.65	0.62	0.63

AUC-ROC Curve

Label	BERT	DistilBERT	Universal Sentence Encoder	Sentence Transformer
AUC-ROC	0.82	0.78	0.76	0.78

Metrics	Weighted F1 Score	AUC-ROC score
Models		
TFIDF + Naïve Bayes Clf	0.58	0.7
W2V + Multi Class Clf	0.58	0.6
BERT DNN Classifier	0.56	0.8
DistilBERT DNN Classifier	0.55	0.78
Universal Sentence Encoder DNN Classifier	0.54	0.76
Sentence Encoder DNN Classifier	0.63	0.78