# Abstract

Over the past decade, admissions rates to selective universities have been significantly dropping, and with so many students having similar academic achievements there has been a greater emphasis placed on the college essay. We examined the correlation between admittance status and college essays by creating a non-biased AI model capable of predicting a student’s admittance to university only based on their essay with 76% accuracy. In addition, we ran sentiment, emotion, and topic analysis natural language processing models to reveal the common characteristics found in admitted essays. The results concluded that admissions officers favored essays with either a strong positive or negative sentiment, that conveyed a strong sense of desire, and were related to an educational topic. These results are an important tool for students and families to gain a better insight into the increasingly confusing university admissions process.

# Introduction

**Background**

Every year tens of thousands of essays are passed into the admissions offices of elite universities. With admissions officers spending on average 11 minutes on each applicant’s file (Harris, 2018), it is difficult for an applicant to make an impression. However, the essays provide the applicants with a chance to connect to their admissions officers to make a greater impression. In reference to the importance of a personal statement essay, Mary Hale Tolar, Former Deputy Executive Secretary of the Truman Scholarship Foundation said, “it is the heart of your application” ([Taylor](https://www.loyola.edu/department/national-fellowships/resources/personal-statement#:~:text=A%20personal%20statement%20is%20your,the%20heart%20of%20your%20application.), n.d.). This shows that admissions officers use essays as an important tool to learn more about the personal characteristics of applicants who often look the same on paper. With roughly 57,000 applicants and only 2,000 spots available for the Harvard Class of 2027 (Havard, n.d), the necessity for applicants to differentiate themselves from others with similar admissions files through essay writing cannot be overstated. However, the ambiguity of what makes a college essay captivating makes it difficult to write an engaging essay. For students, it is hard to determine the strengths of an essay without knowing what is desired from admissions officers., Advances in natural language processing (NLP) technology provide a potential solution, as they now make it possible for machines to analyze large amounts of text data and measure various elements like sentiment and topic while determining the important elements of a piece of text (e.g., Brown et al., 2020). We aim to use these advances in NLP to help students understand how to improve their essays to gain admission to college.

Despite the perceived importance of the essay, the essay isn’t the only factor considered when admitting an applicant to a school. For example, an applicant can have a perfect essay but bad grades, extracurricular activities, and teacher recommendations, all of which are important considerations for admissions officers. We thus first aimed to develop a classification model capable of predicting the success of an applicant based only on the text of the essay. This was an important step to test the theory that there is a correlation between the strength of an essay and the success of the applicant. This proof of concept also validates the importance of essays in the admissions process. This then justifies further investigation into analyzing what factors of college essays increase admissions rates via NLP. In particular, after training an algorithm to predict the success of an essay directly from the text, we further sought to determine what sentiments and topics are favorable in college essays. To do so, we used pre-trained NLP models to extract the sentiment, emotion, and topics associated with each essay and used these to predict admissions results, highlighting which of these factors most aid in admission. In the remainder of the article, we discuss the methods used in further detail, highlight the results of the different models, and provide general advice to students based on our findings.

The results of this investigation also lay the foundation for further research into developing AI tools to assist students in editing college essays. Students could seamlessly input their essays into an AI tool capable of providing real-time feedback and a predictive admission chance for students.

# Methods

Our first objective was to develop a classification model capable of determining if an essay was “good” or “bad”. A good essay would be one that had successful admittance to elite universities. An essay was deemed “bad” if the applicant was denied from universities. Furthermore, essays that were deemed to be bad due to grammatical errors but were not necessarily submitted to colleges were also included to simplify the data collection – see section 2.2 for further clarification. The goal of the classifier was to determine if there was a correlation between the strength of an essay and the success of the application without looking at other factors such as grades and extracurriculars.

The second objective was to use embedding and transfer learning techniques for sentiment, emotional feelings, and topic analysis. The purpose of this was to determine what characteristics were consistent in “good” essays. This provides greater context to the content of “good” essays. We were uninterested in looking specifically at the grammatical difference between essays due to the wide range of sources and tools available for grammatical analysis and assistance. Programs like Grammarly provide on the spot feedback and corrections for grammatical mistakes. However, they do not provide deeper context to the subject of the writing to appeal to admissions readers. We sought to fill that hole in the AI editing sector by finding the stylistic features consistent in “good” essays.

*Data Set for Classification*

In order to train the classifier model we needed to create a data set containing sample essays that were labeled as admitted (good) or denied (bad). We included responses to a variety of different questions and to different Universities to construct this dataset, as many universities require students to write specific supplemental essays as a part of the application process. [Supplemental](https://admissions.upenn.edu/how-to-apply/what-penn-looks-for/essays) essays often deal with questions related to major choice, personal goals, personal growth, service, or why an applicant desires to attend their institution. While not having a standardized question that the samples were answering did influence the topics and style of writing in the response, it also resulted in an easier data collection process. Since there was not an online database of past admitted and denied essays from institutions that we could find, we had to source each essay individually. Many of the admitted essays were easily sourced from college websites that posted successful essays from the past. However, there was a lack of failed/denied essays posted online. People are less likely to post failed work than they are to post successful work. This resulted in difficulty finding essays that fit the “denied” category in the classifier. Creative methods were often used to source these essays. YouTube videos, help forums, and online editing services were all used to source “bad” essays. However, by not limiting ourselves to one response question we resulted in a larger sample size of “bad” essays, resulting in a final dataset of 56 “good” essays and 27 “bad” essays used to train the classifier.

*Data Set for Sentiment, Emotion, and Topic Analysis*

In order to determine which features are most predictive of “good” essays we created a separate database of only accepted essays. For this set of data, all essays were in response to one of the 6 personal statement questions given by the Common App. The Common App is an application format accepted by the majority of universities. The Common App requires students to respond to one of the six prompts listed below in 650 or fewer words.

Common App Questions

1. Some students have a background, identity, interest, or talent that is so meaningful they believe their application would be incomplete without it. If this sounds like you, then please share your story.
2. The lessons we take from obstacles we encounter can be fundamental to later success. Recount a time when you faced a challenge, setback, or failure. How did it affect you, and what did you learn from the experience?
3. Reflect on a time when you questioned or challenged a belief or idea. What prompted your thinking? What was the outcome?
4. Reflect on something that someone has done for you that has made you happy or thankful in a surprising way. How has this gratitude affected or motivated you?
5. Discuss an accomplishment, event, or realization that sparked a period of personal growth and a new understanding of yourself or others.
6. Describe a topic, idea, or concept you find so engaging that it makes you lose track of time. Why does it captivate you? What or who do you turn to when you want to learn more?
7. Share an essay on any topic of your choice. It can be one you've already written, one that responds to a different prompt, or one of your own design.

The standardization of questions for the dataset allowed for more consistent results for the analysis models. All responses to a common app question are regarded as a “personal statement” as is used by universities to learn more about the personal qualities of an applicant. The consistent purpose of the questions allowed for consistent results amongst the models. To collect personal statement essays we sourced samples from college websites again. All essays were from accepted students who agreed to have their essays posted on their college’s site.

To build a classifier we built upon modern advances in natural Language Processing technology. The classifier was built on top of the DistilBert transformer-based language learning model built to do next-word prediction (Sanh & Chaumond & Wolf, 2019). DistelBert is a smaller, lighter, and faster version of the Bert Transformer. This model develops wording embedding representations of the words in the data. Word embedding is an essential technique in the natural language processing field. Individual words get represented as a vector that is then mapped to other vectors with similar characteristics. This opens the gate for text to be used and processed by computers for a variety of different tasks.

We ran a logistic regression on top of the embeddings to perform classification, using the logistic regression implementation from the sklearn python library (Pedregosa et al., 2012). The logistic regression model was chosen after testing the accuracy of several other models from sklearn, including a random forest classifier, a support vector classifier, and a kneighbors classifier. To test the accuracy of the models we used cross validation where the data was split into different test groups and the accuracy from the groups was averaged. This prevented our models from overfitting and providing higher accuracy on training data than on the test data. Through the cross validation of the models, the Logistic Regression model was deemed to have the highest accuracy.

*Feature Extraction - Sentiment, topic, and emotion*

In order to complete the feature extraction of sentiment, emotion, and topic we used pre-trained language models. This allowed for the efficient testing of several models without needing to write and train models from scratch. However, the models were not trained with the intent of being used on college essays. This resulted in some less than ideal results that were later accounted for and fixed by modifying the specification of the models.

For extracting the sentiment we used a model called bert-base-multilingual-uncased-sentiment model, which is based on the BERT architecture (Sanh et al., 2019). The model was trained on 150,000 English product reviews. The model outputs the sentiment of an essay through a representation of 1-5 stars. One meaning the most negative, 3 neutral, and 5 the most positive. We recorded the results in two different formats. One format was of a softmax being calculated over the results. This normalized the probabilistic outcomes of the outputs. However, I was interested in examining the total amount of essays that were deemed to have each sentiment. To solve this problem a relative frequency of argmaxed outputs was recorded too. The argmax determined for each essay what the highest probability of the sentiment was and then recorded that essay for that sentiment. For example, if the highest activated output neuron was the 4 star neuron that essay would be recorded as a 4. Then the relative frequencies of all the stars were calculated and recorded. Furthermore, an analysis was run on the rejected essays as well as the accepted essays.

For extracting the topic of an essay we used a model called distilbert-base-uncased-go-emotions-student which is based on the BERT architecture. It was developed by PhD student and AI researcher, Joe Davison. The model is a zero-shot classifier trained on the GoEmotions dataset, the largest manually annotated data set (Demszy et al., 2020). The model was trained to classify 27 different emotions. However, we only ran the model to classify a total of 21 emotions. This was because some of the emotions were not relevant to our objective of correlating emotions to college essays. The following 6 emotions were removed from the model: confusion, approval, surprise, annoyance, admiration, and disgust. A softmax was also applied to normalize the probabilistic outcomes. The model was run on both the pre and edited model to compare the results.

Finally, a similar approach was used to extract the topic of the essays. The model used was a Website Classifier developed by Ali Mazhar Luqmani. The model is based on the BERT architecture and is a fine-tuned version of the distilbert-base-uncased model trained on an unknown dataset. With over 4.3 million downloads, the model is highly regarded. Similarly to the emotion model, there were many output categories that did not suit our case. The original model was trained to output a total of 16 categories while our version only had 5. The following 11 were removed - adult, social networking & messaging, food, photography, games, e-commerce, forums, streaming services, news, law and government, and business/corporate. The same process was then used of softmaxing the results and comparing the results against the pre and edited model. The results of editing the model were significant.

# Results

*Classification*

The logistic regression classifier had an accuracy of 75.6%, meaning that 75.6% of the time it correctly predicted if the applicant was accepted based on the essay. 67% of the essays in the database were admitted samples. Therefore, the model performed significantly above chance.

*Sentiment*

The sentiment analysis concluded that strongly positive and strongly negative essays were more common in accepted essays. View Fig[1] argmax and Fig[2] softmax for results. Furthermore after testing the rejected essays 5 stars was the most common. Fig[3]

Fig[1]

**Figure 1. Results** - Measures the proportion of essays being rated a given star from 1-5. Scales goes from most negative (1 star) to most positive (5 stars)

Fig[2]

**Figure 2. Results -** Measure the probability of an essay being rated a given star from 1-5. Scales goes from most negative ( 1 star) to most positive ( 5 stars) with 3 stars being neutral.

Fig[3]



**Figure 3. Results -** Compares the relative frequencies of the number of stars in accepted and denied essays.

*Emotion*

The emotional analysis of the original model concluded that essays evoking a sense of confusion were most commonly found in accepted college essays with the emotions of desire, caring, and approval being closely found to also be successful. View Fig[3]. The edited model concluded that essays evoking a sense of desire and curiosity were very common is successful essays with remorse, curiosity, and excitement were closely found to also be successful. View Fig[4].

**Figure 4. Results -** Shows the probability of a personal statement essay evoking a particular emotion.

Fig[5]

**Figure 5. Results -** Shows the probability of a personal statement essay evoking a particular emotion.

*Topic*

The topic analysis of the original model found education to be the most common topic in accepted essays by far. With the edited model supporting this claim. View Fig[6] and Fig[7].

Fig[6]

**Figure 6. Results -** Shows the probability of a personal statement essay being about a particular topic.

Fig[7]

**Figure 7. Results -** Shows the probability of a personal statement essay being about a particular topic.

# Discussion

*Classification*

The results from the classifier show that artificial intelligence is able to predict the admittance of a student based on nothing but their college essay. Furthermore, this reveals a correlation between admittance at the strength of the essay. With the model correctly predicting the admittance 76% of the time in a set of data where admitted essays were only 67%, this shows a correlation. The results can also be improved with a larger and more consistent dataset. Due to data collecting limitations, the data set was small and involved several different essay prompts for the training of the classifier. A larger and more consistent dataset will allow the model to better learn the nuisance of a prompt and what is important for each prompt. For example, if a large data set of admitted and denied students responded to a prompt regarding community service then the model will be specialized in handling the nuances admissions officers seek in community-related essays.

Furthermore, using AI to predict an applicant’s admittance to a university can be strongly improved if a model were to consider the context of an applicant's application. While important, an essay is only a small piece in the large puzzle of admitting students to a university: grades, extracurriculars, geographic location, socioeconomic status, race, and legacy are all important factors that admissions officers consider when reviewing an application. A future model capable of handling additional inputs of such will have a significantly higher accuracy in predicting students' admittance.

The results of such a model can be crucial for many high school students looking to apply to colleges. Such a model can be used as a tool for students to make decisions about which institutions to apply to based on their calculated chance of admittance. Current admittance calculators only account for factors like GPA and standardized test scores but if accompanied by the student’s essay results can be significantly accurate since the results of this experiment found a correlation between essays and admittance. Furthermore, such an AI model can be used by students to help reveal weak points in their application and help them to improve it.

College admissions is an already very ambiguous process that often feels unpredictable and random. However, now with the use of AI, it is possible to reveal the ambiguities that trouble so many students a year.

*Sentiment, emotion, topic*

The results of the sentiment analysis reveal that admissions officers are looking for strongly negative or positive essays in the personal statement. This is not surprising as personal statements often follow the structure of a personal narrative, and a narrative without strong sentiment can be boring. This is backed up by desire, caring, and remorse being the three most common emotions present in accepted personal statements. This shows that officers are looking for strong and powerful stories highlighting the author’s motivations and passions. In addition, according to the topic analysis, it is important for an essay to be related back to education. Authors need to show admissions officers that they value their education.

In addition, the essays from the denied sample surprisingly overwhelmingly had a positive sentiment Fig[3] with about 50% of the essays being rated 5 stars. I expected the denied sample to have a mostly neutral sentiment (3 stars). However, these results are unreliable. The dataset only contained 27 samples. In addition, due to sourcing limitations, they were in response to a variety of different supplemental prompts from different schools. Whereas the essays used in the accepted essay sentiment analysis were all in response to a personal statement prompt. Since the sample data for the denied essays in the sentiment analysis were far fewer than the samples for the accepted essays, and the denied essays were not consistently answering the same question, the results of this trial should not be heavily weighted.

My advice to students based on the results of this study is to write from their hearts and be honest with themselves. It is evident that admissions officers are looking for genuine people who are excited about the university and are going to fully maximize their experience.

The results of this study have removed some of the ambiguity behind the admissions door. While it is still not clear what will get a student accepted, we now know a little more about what is desired from essays.

*Conclusion*

It is now possible to start bridging the gap between college admissions officers and hopeful applicants. The results show that there is a correlation between an applicant’s essay and their admittance. This justifies more future work and research into creating a comprehensive AI tool for commercial purposes. A nationwide independent study undertaken by the research firm Lipman Hearne, aided by the National Research Council on College and University Admissions, found that 26% of high achieving high school students (those seeking a four-year college) used a private college counselor (Lipman Hearne., N.D). Furthermore, the Independent Educational Consultants Association (IECA) reports that many counselors charge just under $140/hour (ICEA,. N.D). With obvious market demand and money flowing into helping high schoolers with the college application process, a comprehensive AI tool that is able to provide real-time feedback and specific suggestions about a student’s college essay would be invaluable. The results of this investigation lay the foundation for further research into developing such tools: students can seamlessly input their essays into a similar tool capable of predicting the chance of that essay getting admitted, altering their essay to optimize the predicted admission chance. Separate models can be trained for different schools and different prompts. Furthermore, a model can also be trained to interpret a student’s grades, extracurricular, and test scores. Such a model would be able to offer significant predictions in regard to a student’s chance to be admitted into a certain university. This could help students save time and money when applying to schools. I am excited to see where future research takes this study.

#

# References

*Admissions Statistics | Harvard*. (n.d.). Harvard College. Retrieved October 9, 2023, from <https://college.harvard.edu/admissions/admissions-statistics>

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Demszky, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A Dataset of

Fine-Grained Emotions. *ArXiv*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

*Essays | Penn Admissions*. (n.d.). Penn Admissions. Retrieved October 17, 2023, from https://admissions.upenn.edu/how-to-apply/what-penn-looks-for/essays

*FAQs About the Independent Educational Consulting Profession –*. (n.d.). Independent Educational Consultants Association. Retrieved October 17, 2023, from https://www.iecaonline.com/quick-links/ieca-news-center/press/background-information-on-independent-educational-consulting/

Hale, M. (n.d.). *What is a Personal Statement - National Fellowships*. Loyola University Maryland. Retrieved August 28, 2023, from https://www.loyola.edu/department/national-fellowships/resources/personal-statement

Harris, Z. (2018, September 6). *A Step-By-Step Guide on How College Admissions Officers Read Your Application*. InGenius Prep. Retrieved October 9, 2023, from <https://ingeniusprep.com/blog/college-admissions-officers/>

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in

Vector Space. *ArXiv. /abs/1301.3781*

OpenAI. (2023). GPT-4 Technical Report. *ArXiv. /abs/2303.08774*

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., …, & Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python. *ArXiv. /abs/1201.0490*

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller,

faster, cheaper and lighter. *ArXiv. /abs/1910.01108*