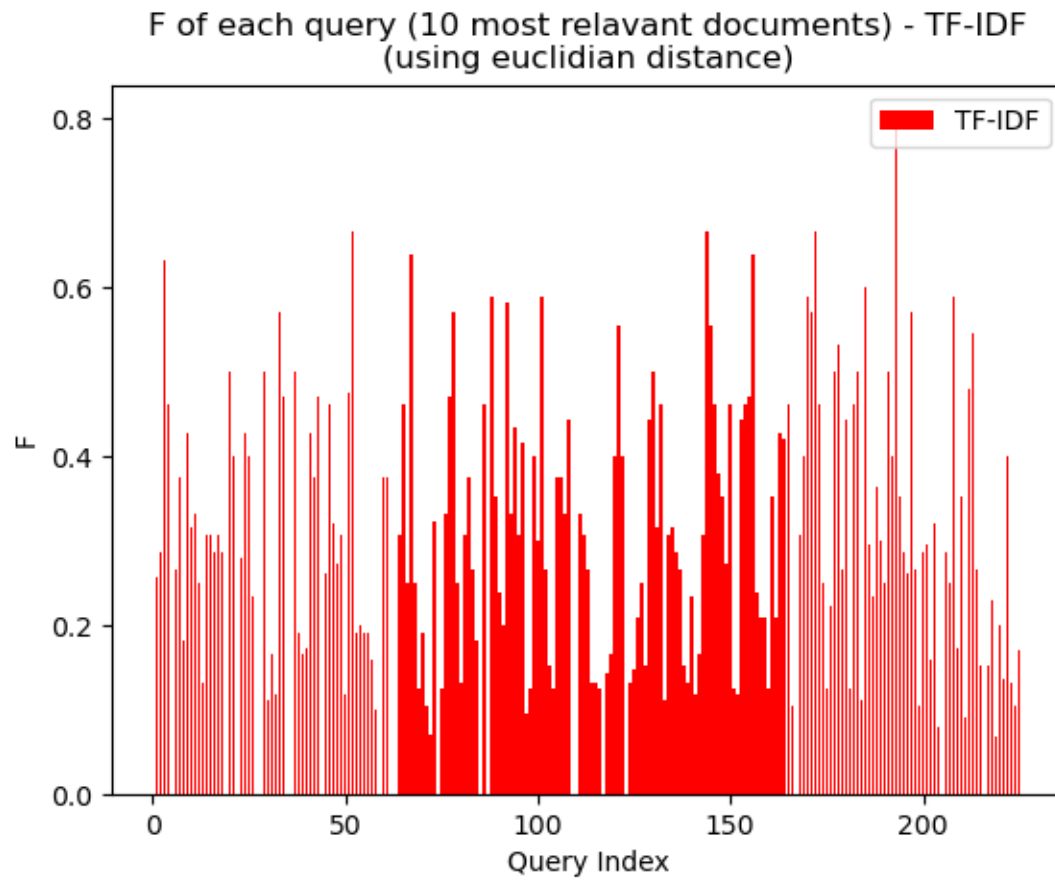


Alena McCain  
U00833503  
[brand.18@wright.edu](mailto:brand.18@wright.edu)





Alena McCain  
U00833503  
[brand.18@wright.edu](mailto:brand.18@wright.edu)

In general, both metrics performed equally well, using the TF IDF model. Cosine similarity performed similarly well when using the binary model. However, Euclidean distance was comparatively poor when being used with the binary model.

I suspect Euclidean distance performed similarly to cosine similarity in the TF IDF model because of the document format. Since all the documents were abstracts, one would expect them to be of similar length. This could be removing one of the major biases related to Euclidean distance's poor performance, the document vectors' length. That is to say that two documents of a similar length might be considered more similar even if the content varies because they have a closer distance than a larger document with similar content. Using abstracts that are similar in length removes this issue.

On the other hand, Euclidean distance may have performed worse than cosine similarity in the binary model because the binary model doesn't account for term frequency. Therefore, while it is useful that the abstracts are all similar in length; to dampen biases in the Euclidean distance, this proves that the term frequency is still relevant when using the Euclidean distance metric. This is because the term count has a more drastic impact on a measure like Euclidean distance, as opposed to cosine similarity. The angle between two vectors will always remain the same regardless of the term frequency in the document, however, the lengths could change.

In general, the TF IDF model performed better than the binary model. This makes sense because including term frequency in metric calculations allows for another means of comparing the documents. If a lot of the abstracts share similar terms, the only way to distinguish them during ranking is by term frequency.

It is also interesting to note that each of the models and metrics tended to perform better on queries approximately between 60 and 160. This is probably because the other queries were very specific or general; high precision and low recall would indicate the former, and the reverse would indicate the latter. However, this sort of consistency between the models and the metrics is useful in getting a benchmark for other potential models, or to determine if one of the models presented here was performing strangely.

That said, the average precision among both the TFIDF metrics and the Cosine Similarity Binary metric were between 25 and 30 percent, which seems low, compared to a standard search engine like Google. Recall was much higher, between 35 and 40 percent. This may be because of the number of queries that had less than 10 relevant documents, meaning it's easier to achieve a recall of 100 percent. I hesitate to put much emphasis on the recall because some queries had a lot more than 10 relevant documents, which would lower the average recall by no fault of the model. This is also evident from looking at the F scores of the three models which are between 26 and 30 percent. Since the F scores are closer to the precision, it is probably true that the recall is affected by the number of relevant documents for each query, as opposed to being an accurate reflection of model performance.