**Rubric**

Overall themes (10%):

* A clear statement of the problem you are trying to solve.

Detecting transcript anomalies that could indicate the company is about to be engaged in a M&A transaction

* Novelty of your approach to the problem.

Most M&A related NLP work centers around using annual or quarterly filings (10K/10Qs) or new articles. We are using company earnings and conference transcripts.

* Who else has worked on this problem? What did they do? What are you doing differently?
  + [Predicting merger targets and acquirer from text abstract](https://www.aeaweb.org/conference/2017/preliminary/paper/Z5dzeKEd) : This paper use only a portion of 10k/10Q disclosure.
  + [Using ML to analyze M&A](http://giovanniperi.ucdavis.edu/uploads/5/6/8/2/56826033/tiffany_jiang.pdf) (use only 10k data but more sections)
  + [Deal or No deal: predicting M&A at scale](https://ieeexplore.ieee.org/document/9006015) (also 10k data only)
  + [Text Analysis on M&A](https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1014&context=studentpub_uht) (less relevant, it is an analysis of M&A documents)

* Is it clear you read and understood the papers you cited? Are they the right ones?

Crisp objective (20%):

* "How will I know when my project is successful?", usually in the form of an evaluation metric.

AUC score?

* Tied back to overall problem statement - why is this the right objective?

The objective is be able to rightly classify those transcripts that points to insider knowledge

* Statement of a clear baseline (e.g. for classification, predict most common class for everything).
* Is this the objective others working on this or similar problems use? If not, why?

Methodology and analysis (50%):

* Are you using appropriate techniques for the problem you're trying to solve?
* What did you learn from the first approaches you made? How did you take what you learned and use that to improve subsequent iterations?
* What weird patterns in the data have you found? What about how the model interacts with your data? (Did you find any interesting loss patterns?)
* Do the patterns you observe align well with patterns other researchers have found in the papers you read? What did they do about them?

Technical communication (20%):

* Primarily:
  + Succinct, interesting presentation explaining the above.
  + Is your paper well organized?
* Table stakes:
  + Did you run spell check before submitting?
  + Did you proof read it?

Volume of work (built into the other components):

* Is this someone else's code (downloaded from github) simply run on already cleaned up data (downloaded from Kaggle)?
* Did you run assignment code on a new dataset and report just the analysis we built in the notebooks?
* Especially for teams of 3, is it clear each person did a solo-project amount of work?