Apr 23rd, 9:00 AM

# Using data mining to analyze job reviews

Nicholas Bornkamp
*Rowan University*

Tony Breitzman
*Rowan University*

Rowan University
Department of Computer Science

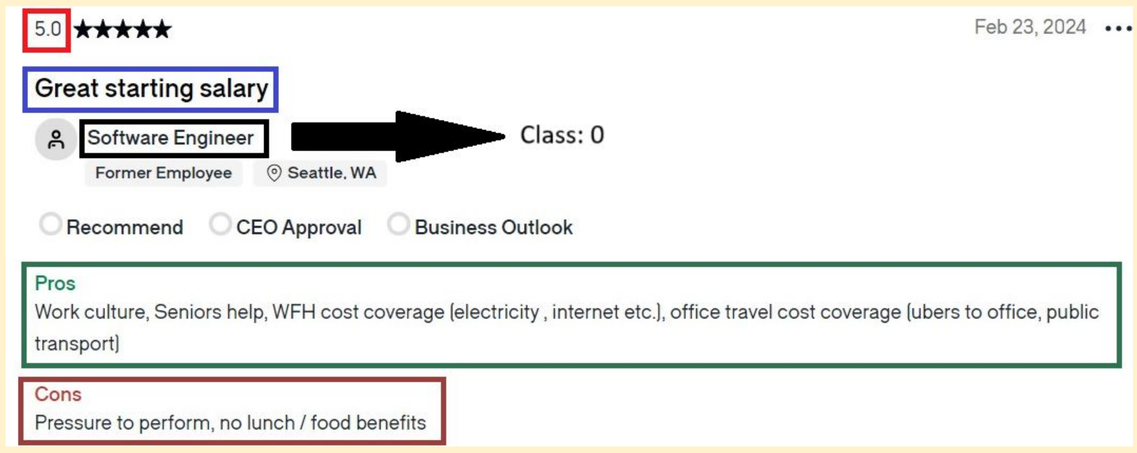
# Using Data Mining to Analyze Job Reviews

**Nicholas Bornkamp**
PI: Anthony Breitzman

## Introduction

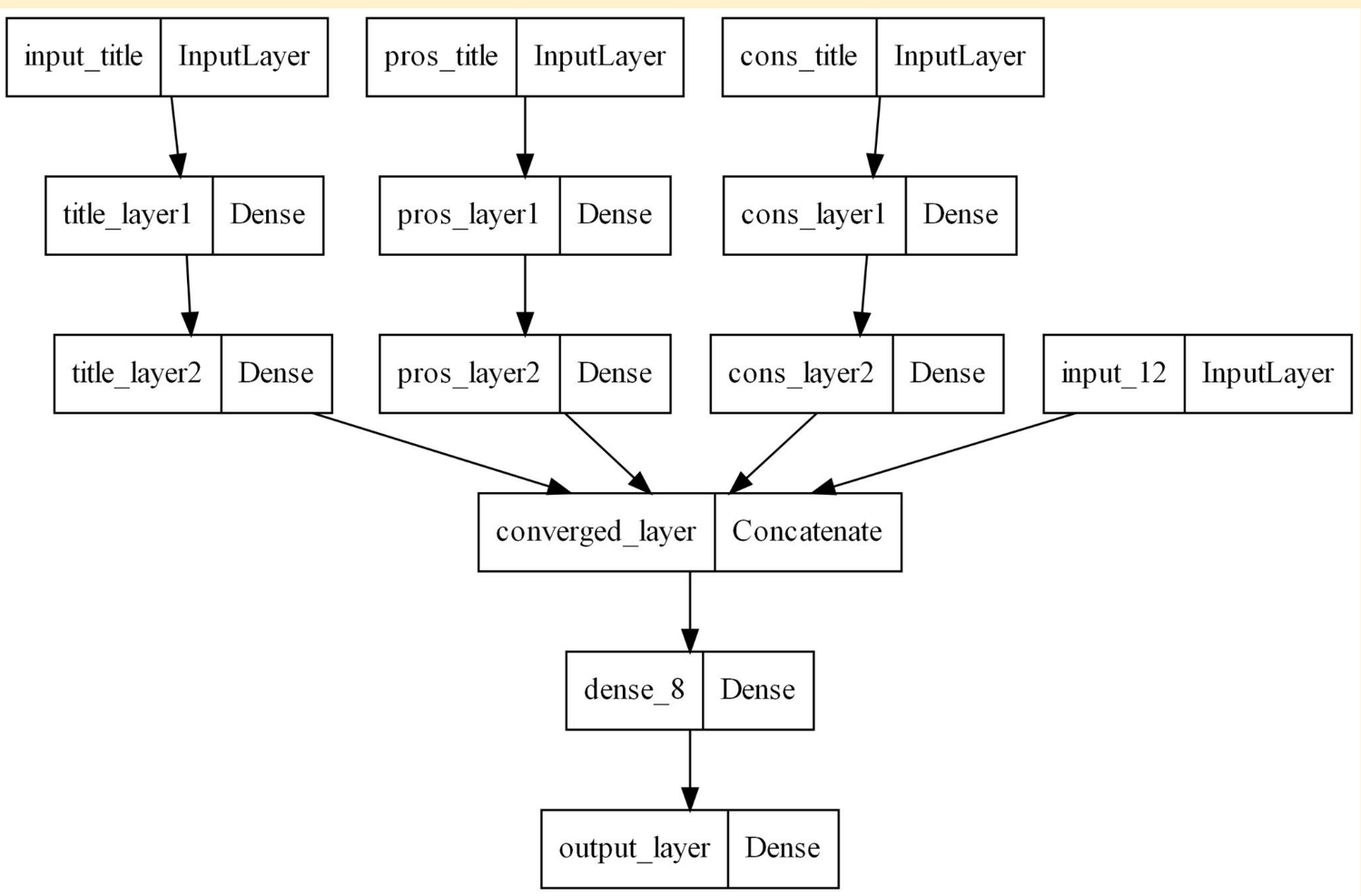
Job review websites like Glassdoor are not always clear on how well the company operates, especially as viewed from differing levels of employment. For instance, a middle or upper manager from Amazon may have an overall positive review of the company with minor issues about it, but someone who works in the warehouse may have a mixed experience. To solve this issue and determine any correlation between employee level and their review, data mining techniques were utilized such as website scraping and neural network training to develop a model that analyzes employee reviews.

## Data Scraping and Parsing



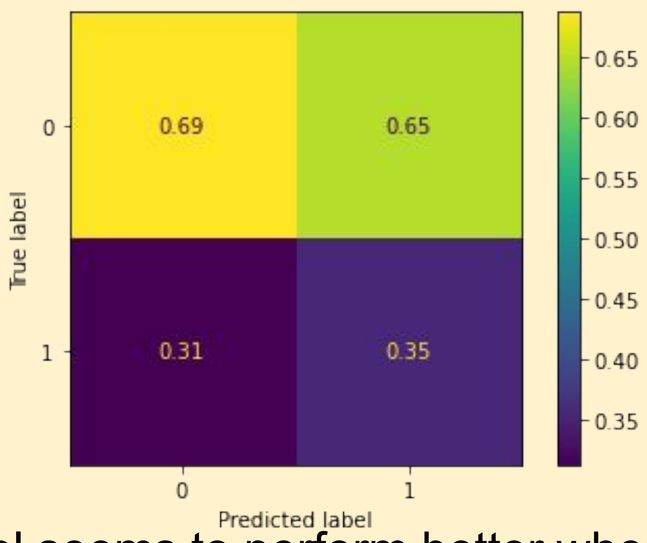

- About 5,000 reviews were scraped from companies such as Amazon, Tesla, and Microsoft.
- The Review Title, Pros, and Cons were analyzed with nltk SentimentIntensityAnalyzer to calculate the sentiment for each review part.

## Model Structure




- Multiple input layers are used to focus learning on the three parts of the review as well as an additional review score. The sentiments go through the network and are concatenated at the end along with the review score (scaled from 0 to 1).
- A final Dropout layer is used to help the model generalize better while it is trained on less data.
- The output layer uses a sigmoid activation to match the binary classification problem, where 1 represents an employer/upper level worker (i.e. manager, executive, etc.)

## Results and Analysis




- The model seems to perform better when identifying employees versus employers, which could be a result of oversampling from the 0 class.

## Potential Improvements

- Implementing Tokenizers for the wording of Pros and Cons, in addition to developing a NLP model with LSTM layers, could prove effective in analyzing key phrases from reviews.
- Rather than utilizing an ensemble method approach, a single Input layer could be used and split into multiple branches.
- More companies can be scraped, and more reviews per company can increase the dataset size and allow for more diverse data.