AIT526
Introduction to Natural Language Processing
Term Project - Final Report
Sentiment Analysis on RateMyProfessors's Reviews

1. Title:

Proposal Title	Sentiment analysis on RateMyProfessor reviews.
Team	Team 6 Ashish Pandey Hadeel Elyazori Kirtana Nambiar Siddharth Nandargi
Course Project Professor's Name	Dr. Kevin Lybarger
Course Name	AIT-526-001 Introduction to Natural Language Processing
University Name	George Mason University
Date	11/02/22

2. Abstract and Keywords:

Class registration is an important procedure that university students go through every semester to decide which courses they will register for and with which professor. One of the techniques students use to assist them with that is RateMyProfessors platform, where other students provide feedback on professors, they took courses with. Part of the students' review is contextual, which is their comment on the specific professor. The task that was performed in this project is using the power of Natural Language Processing (NLP) at analyzing human language, to perform sentiment analysis on the text. This will be done to extract the sentiment of the text and classify it into positive, negative, and neutral. The beneficiaries of this analysis are both the students and professors. While the students will be able to decide on whether to take a class with a certain professor, it will also help professors pinpoint the points that upset the students and the ones which satisfy them.

This task was performed using supervised machine learning. The models that were trained and utilized are K-nearest Neighbors (KNN), Logistic Regression, Random Forest and Recurrent Neural Network (RNN). They were used to predict the sentiment of the comment in the review and their results were compared. Moreover, topic modeling was performed for further analysis.

Keywords: RateMyProfessors, Natural Language Processing (NLP), sentiment analysis, K-nearest Neighbors (KNN), Logistic Regression, Random Forest, Recurrent Neural Network (RNN), topic modeling.

3. Introduction:

Sentiment Analysis is the process of extracting subjective information like emotion, underlying in each text. It is widely used in identifying overall opinions of mass towards a particular topic, and this insight then helps business or governments to respond appropriately to meet their user's needs. With the boom in data in today's era, everyone today makes it a point to read reviews before visiting a restaurant or buying a product to arrive at a decision. Hence, leveraging this information to make any experience utilitarian would not only benefit the customers but also the business to improve their products and satisfy customer requirements.

In this project, we attempt to perform sentiment analysis on RateMyProfessors website which contains reviews given by students, in order to get a summarized sentiment on the different teaching methods of professors. This analysis will help the professors to get brief feedback without having to go through all the countless reviews and important keywords and help them take necessary actions to make the learning experience a holistic one. It will also provide the students with details about the teaching methods or any attributes of the professor and help them in the decision to pursue a class as per their preference of learning. The main aim of the project is to perform sentiment analysis to categorize the professors' teaching methods into positive, negative, or neutral classes. Further, utilizing these polarity scores, we intend to train a machine learning model to make accurate predictions of sentiment based on polarity obtained from reviews, star rating, whether the students would take the class again, tags given to each professor, difficulty index of class, etc. Apart from this, extracting keywords associated with positive or negative reviews using topic modelling would help us identify some of the key components to address the issues related to learning and thereby helping professors and students in attaining a universal learning experience.

4. Related Work:

Sentiment Analysis for RMP has been performed previously but the scope is limited to either finding out sentiments from the posted reviews, and the dataset size is restricted to fewer professors from a single university. Past work on this topic has primarily examined usefulness and sentiment of these ratings (Otto, Jr, and Ross 2008). In one of the previous work, focus was more on correlation levels between easiness, clarity, and helpfulness of the professors (Otto, Sanford, and Wagner 2011). In a more recent study, which advances the ones mentioned earlier, the extent of the project is such that it categorizes and examines Professor reviews based on the tier of universities and yet again has performed an analysis considering only 75 universities in the USA (Ziqi Tang, Yutong Wang, Jiebo Luo, 2021).

Rich data available on the RMP (RateMyProfessor) website provides a wide breadth for mining numerous insights about professors' teaching methodologies, characteristics that influence the quality of teaching, information on characteristics that are considered important by students while posting reviews for professors, etc. which can really help in uncovering useful insights to understand and improve the learning experience. To achieve this, ideally, the best approach to the solution will be to use in conjunction with NLP (Natural Language Processing) concepts like sentiment analysis and performing EDA (Exploratory Data Analysis) which will be helpful in

creating and tuning ML (Machine Learning) models for classification, and topic modeling to identify the keywords. Areas where these methodologies fall short would be due to some bias in the datasets. As observed, most of the reviews are posted in the months of May and December, which are the ends of the semester. Also, the data does not describe which courses have transformed from online to in-person classes after the COVID-19 situation was diminished. These NLP tasks mentioned would produce sentiment results without considering any of these constraints. The above-mentioned concepts/approaches align with our proposed project work as sentiment analysis helps in generating a summary of the different teaching methods of professors and help in categorizing into positive, negative, or neutral class. Polarity scores will aid in training a machine learning model to make accurate predictions of the sentiment considering the numerous variables present in the dataset, while topic modeling will help to extract keywords associated with positive or negative reviews and understand the factors influencing the quality of study or factors which students consider of priority while posting comments.

Use of the three NLP tasks mentioned above in confluence, makes our project work different from the similar work being carried out on this topic as the scope of this work is not restricted to generating just the polarity scores for sentiment analysis for a limited reviews from specific universities, but to understand the key components to address the issues related to learning and thereby helping professors and students in attaining a universal learning experience on a large number of ratings from numerous universities.

5. Objectives:

- Give the students a general insight about the professor before dwelling on details or reading through the reviews, to help them shortlist their choices and then go through the reviews thoroughly to decide to take the class with the respective professor.
- Give students insights about the techniques the professor uses in his/her teaching, and his/her methodology in handling the course.
- Give the student the average ratings on some of the traits of the professor through keywords (which is one of our features) used by other students. For example, it can tell the student that 80% of the students added “Amazing lectures” to their review of this specific professor. Another example is when the student is only interested in getting a high grade to raise his or her GPA, irrespective of other qualities of the class and/or professor, the system will be able to provide the percentage of students who classified the specific professor as “Tough grader” or “Easy grader” through the keywords.
- Give the professors the ability to have an overview of the students’ feedback on them, and insight into whether the overall reviews is negative or positive.
- Give the professors keywords to help them in tackling the issues students are reporting and fixing them, such as “Tough grader” or “Gives bad feedback.”

6. Dataset:

The dataset we are using for our project is shared by a professor at Tsinghua University, called Dr. Jibo He (He, 2020). The data in it was obtained by crawling RateMyProfessor.com which is a

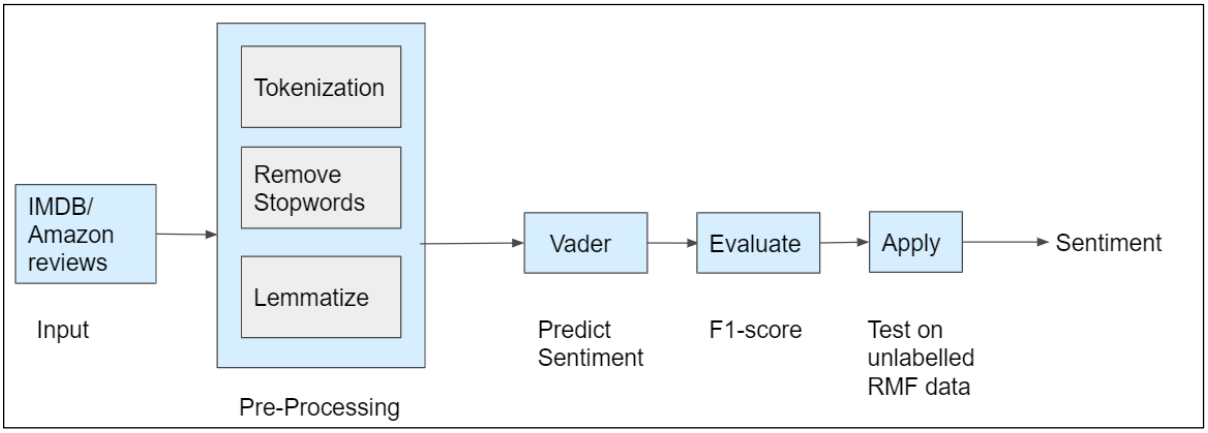
website for evaluating professors by students, and which is the website we are aiming at analyzing. It has data crawled that includes almost one million professors' information and is stored in a .csv file which has 18 variables.

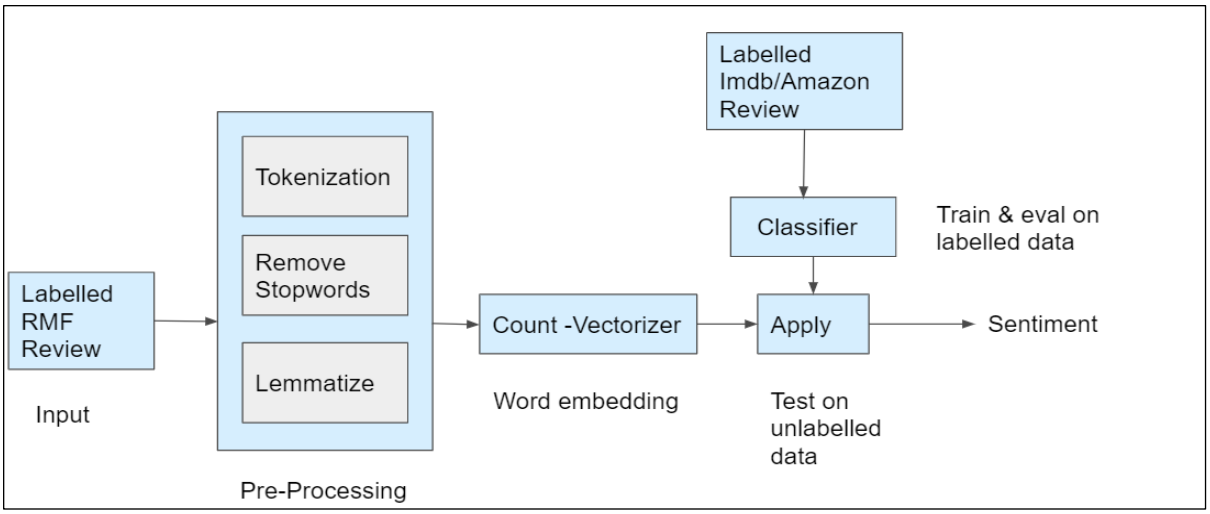
The aforementioned dataset, which we will be using, is composed of 52 columns and we are interested in several columns for our project:

- Professor's name: gives the name of the professor.
- School's name: the name of the university where the professor teaches.
- Star rating: his average star rating score.
- Take again: the percentage of the students who are willing to enroll in a course with this professor again.
- Tag professor: the keywords the student used in describing the professor, such as "good grader", and "gives good feedback."
- Attendance: whether a course is mandatory or not
- Difficulty index: the level of difficulty of the course.
- Comments: the comments given by the students about the course and the professor
- Grades: the grade the student obtained for the specific course with the specific professor

7. The System:

Architecture/framework





The NLP and data analytics approaches used to work on the problem are:

1. **Classification:**
   - **Logistic Regression:** Since Logistic Regression uses the sigmoid function to produce a probability between 0 and 1, they are also used to forecast the sentiment based on the features that have been extracted.
   - **K-Nearest Neighbour:** As sentiment analysis is a binary classification, it is a good practice to consider KNN as a classification model
2. **RNN:**

   As we tokenize the sentences, we can parse tree of each sentence into a single input by Recurrent neural Network, which is processed independently. There are multiple ways of implementing RNN, they're – single long short-term memory, bidirectional long short-term memory etc.

3. **Rule Based VADER models:**

   VADER is a rule-based approach which is present in nltk package. Vader sentiment analysis maps context to the lingual features and calculates the scores (polarity). Scores are obtained by summing up the severity of the words in the reviews or comments. The output gives the scores for all sentiments including positive, neutral, negative and compound.

4. **Topic Modelling:**

   Topic modelling is a statistical approach to get the abstract that occur in a collection of documents. We used the Latent Dirichlet Allocation for building topic models where we obtained the keywords and association of words using which we developed a word matrix. We then applied this topic model on our comments/reviews dataset for each professor to obtain a gist/summary of the numerous reviews posted for a specific professor. Grouping and clustering techniques helps us build topics based on keywords and saves time by no reading each of the reviews posted by students.

The software development platforms that were used are:

1. **Jupyter Notebook:**

   Jupyter Notebook can be used to explore the data through visualization. Data visualization can be done using various python libraries.

2. **Google Colab:**

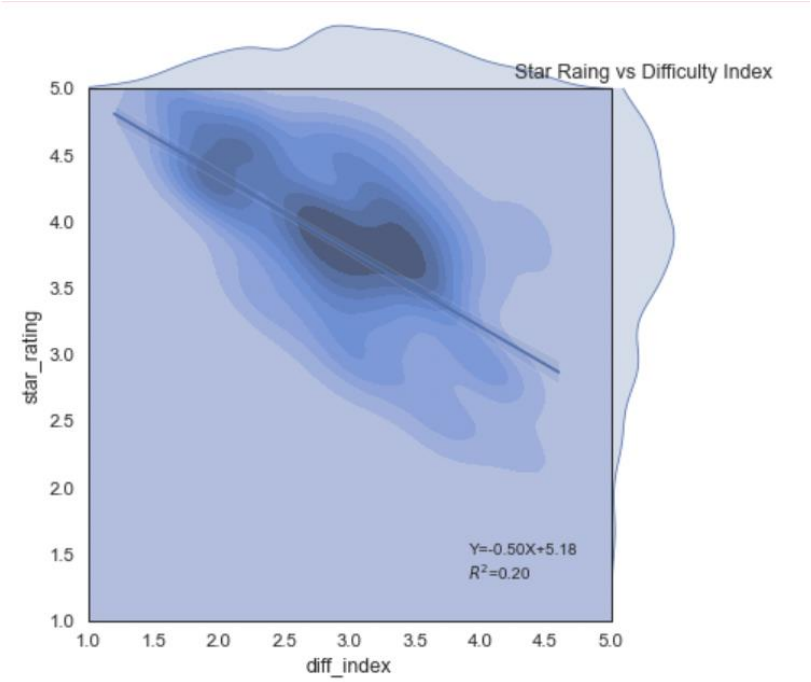
   Google colab will be used for preprocessing the data, building and model and evaluating it. As it is a cloud-based platform, we can install any number of packages or libraries which are not compatible in the local system. Example, Google colab will make spaCy run easily while local system might not be able to run it.

The hardware development platforms that were used are:

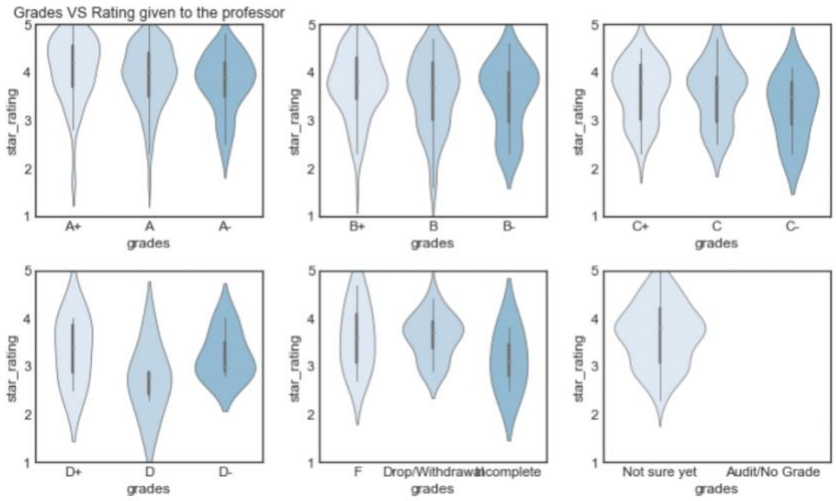
No direct hardware interfaces were involved except the machine having basic processor, RAM and hard disk.

CPU	8 cores with 4 performance cores
GPU	10 core GPU
Neural Engine	16 core neural engine
Bandwidth	100GB/s memory bandwidth

8. Experimental Results and Analysis:



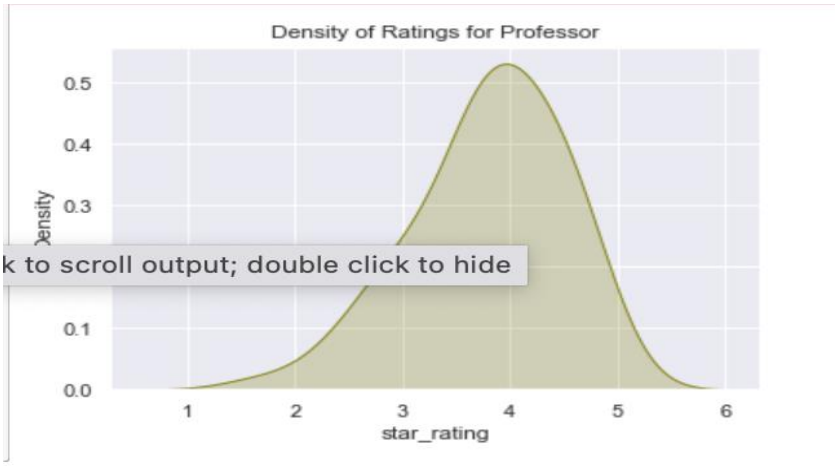
An interesting insight, we found that the star rating of the professors was good when the difficult index of the course was less, as the difficulty level eventually increases the rating of the professors.



The above insight shows the ratings given to professor based on the grades obtained by the student. It shows that, people with grade A have given good ratings to the professors, it's obvious that there can be more factors not just the top grade.

	high_professor_tag_count
Gives good feedback	254
Respected	254
Caring	226
Tough grader	212
Skip class? you won\'t pass.	191
Participation matters	185
Inspirational	179
Amazing lectures	177
Clear grading criteria	174
Get ready to read	172
Hilarious	168
Lecture heavy	140
Lots of homework	126
Accessible outside class	104
Extra credit	97
Graded by few things	67
Test heavy	58
Group projects	55
So many papers	48
Beware of pop quizzes	46

This insight will show the tags used to the professors with high ratings.



The above insight shows the rating density of professors ranging from 1 to 5 and shows the peak at 3.8-4.

• Analyze and interpret the results

Algorithm	**F1-score**
KNN	0.63
Vader	0.73
Logistic Regression	0.78
Random Forest	0.83
RNN	0.90

As seen above from the table, RNN performs the best as compared to the rest.

The proposed solution outperformed the Vader Sentiment Analysis (Baseline model) and this could be because of the technique used in RNN to train the model as every layer output is remembered and taken as input to next layer, thereby remembering historical information which is extremely important in sentence understanding and mimics a human being.

Eg: Classifying a sentence as positive or negative using RNN trained model.

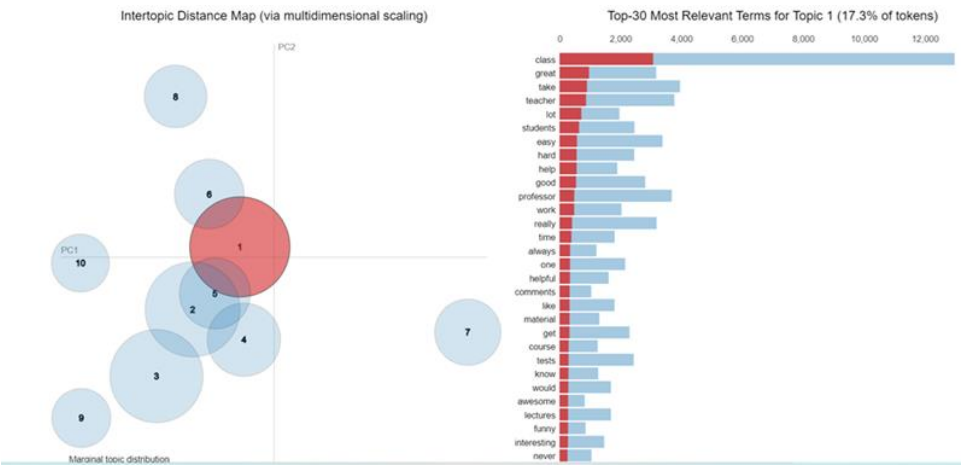
```
##predicting sample sentence sentiments

sequence = tokenizer.texts_to_sequences(['Excellent professor. Hilarious, fun, and good.'])
test = pad_sequences(sequence, maxlen=max_len)
pred = model.predict(test)
if pred > 0.5:
  print('Positive')
else:
  print('Negative')

1/1 [==============================] - 0s 34ms/step
Positive
```

• Results for Topic modelling using LDA

- Log likelihood and model perplexity were considered as metrics to assess the performance of the implemented topic models. Log likelihood score is –235154.6 while model perplexity is 383.09. Higher log likelihood and lower model perplexity are indicators of a better performing model.



Topic model clusters created using pyLDAvis python library where each circle represents the topics based on the frequent word tokens present in the comments/reviews.

```
Topics = [" Punctual, Engaging Class and Helpful Professor.",
 "Lots of Class test/exams,teacher helps with homework when asked.",
 "Helpful professor.Fun learning class.Provides enough review/study material.",
 "Does not listen to the world. Avoid this teacher.",
 "Professor loves his subject. Makes time to talk after class. Lot of knowledge from work experience.",
 "Regret taking this professor's class. Weekly tests makes it difficult to pass his class.",
 "Lot of paper reading and assigns grade for that. Students in his class fail.",
 "Tough time understanding accent. Boring lectures about research.",
 "Professor recommended. Enjoyed his class and teaching style.",
 "Professor takes care of all students. Always attend, never miss. Coursework helpful."]
df_topic_keywords["Topics"]=Topics
df_topic_keywords
```

Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10	Word 11	Word 12	Word 13	Word 14	Topics
come	year	college	require	way	mind	minute	day	provide	suggest	stay	engage	favorite	board	Punctual, Engaging Class and Helpful Professor.
test	question	study	book	exam	read	time	grade	quiz	teacher	answer	ask	help	homework	Lots of Class test/exams,teacher helps with ho...
make	understand	try	math	professor	fun	help	student	teach	feel	material	review	time	explain	Helpful professor.Fun learning class.Provides ...

Keyword Document matrix and the list of topics based on them.

9. Conclusions:

- We were able to predict the overall sentiment associated with each professor. This was used or will be used to summarize the sentiment of students on the different teaching methods of the professors. When this system is used, it can help both the students and the professors to take the necessary actions to make the learning experience a holistic one.

- We also performed topic handling to provide summarized keywords. We did this using the comments of the reviews to generate topics using the LDA (Latent Dirichlet Allocation) model. This can be used to enhance assigning topic keywords to know more about the subject.

Future work:

- Typically, fine tuning the hyper parameters of the models will be something to explore for a better accuracy.
- Working on balancing the dataset using techniques like SMOTE (Synthetic Minority Oversampling Technique)
- Extracting more features to improve model prediction
- Training the models with a domain specific labelled dataset to improve the relevance and accuracy of the models.
- Implementing active learning as an alternative method to label datasets.

10. References:

[1]. He, Jibo (2020), "Big Data Set from RateMyProfessor.com for Professors' Teaching Evaluation", Mendeley Data, V2, 10.17632/fvtfjyvw7d.2

[2]. Otto, Jr, and Ross 2008 "Does RateMyProfessor.com really Rate my professor" Assessment & Evaluation in Higher Education, v33 n4 p355-368 Aug 2008

[3]. Otto, J., Sanford, D., & Wagner, W. (2005). Analysis Of Online Student Ratings of University Faculty. Journal of College Teaching & Learning (TLC), 2(6). https://doi.org/10.19030/tlc.v2i6.1833

[4]. Ziqi Tang, Yutong Wang, Jiebo Luo, 2021. Mining Comments on RateMyProfessor.com https://doi.org/10.48550/arXiv.2101.12339

[5]. He, Y., Xiao, K., & Shi, Y. (2020). Discovering Unknown Reflection. Discovering Unknown Reflection - Sentiment Analysis. Retrieved November 2, 2022, from https://files.eric.ed.gov/fulltext/ED611630

[6]. Sentiment Analysis(ML & RNN). (n.d.). Kaggle.com. Retrieved December 4, 2022, from https://www.kaggle.com/code/smitshah00/sentiment-analysis-ml-rnn

[7]. Brownlee, J. (2020, February 11). Bagging and Random Forest for Imbalanced Classification. Machine Learning Mastery. https://machinelearningmastery.com/bagging-and-random-forest-for-imbalanced-classification/

[8]. Li, S. (2018, June 2). A Beginner's Guide on Sentiment Analysis with RNN. Medium; Towards Data Science. https://towardsdatascience.com/a-beginners-guide-on-sentiment-analysis-with-rnn-9e100627c02e

[9]. Amidi, A., & Amidi, S. (2019). CS 230 - Recurrent Neural Networks Cheatsheet. https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks