## CSCE 771: Computer Processing of Natural Languages

## Prof. Biplav Srivastava, Fall 2022

## Quiz 3 / Instructions

* This is a programming quiz. Code has to be submitted in a directory of your GitHub called “Quiz3” with sub-dir for code, data and doc. Code will have your source code, data will have any input or output generated, and doc will have a .pdf of this file (called Quiz3-CSCE771-answers.pdf) along with any answers
* Complete quiz by 9:00 am on Monday, Nov 21, 2022 by sending an email to [biplav.s@sc.edu](mailto:biplav.s@sc.edu) confirming completing the quiz and attaching your Quiz3-CSCE771-answers.pdf.
* Total points = 70 + 30 = 100
* Obtained =

Student Name: Guangyi Chen

The objective of the Quiz is to learn usage of large language models on NLP tasks and their superiority, if any, over traditional methods. Then, to also solve a practical problem.

**Dataset**: South Carolina universities have “Annual Security and Fire Safety Reports”. For 2022, for University of South Carolina, it is publicly available as well as conveniently cached at: <https://github.com/biplav-s/course-nl-f22/blob/main/sample-code/common-data/2022-uosc-securityandfirelreport-1001bcleryreport.pdf>

**Goals:** Your task is to use NLP techniques to provide specific information to prospective new students and their parents who do not have the background or time to read the document.

**NLP Tasks:** Entity extraction, sentiment mining, events, topic analysis and text summarization

**Activity:**

- Choose any 3 NLP task and corresponding goodness metrics. (You may use additional task for extra credits but mark it so in your report/ code)

- Use any LLM available from Huggingface like BERT, DistilBERT. Use [1] for reference.

- Use any one traditional NLP method (i.e., non-LLM) for the NLP tasks (like extractive summarization based on TF-IDF as discussed in class).

- Now answer the questions and their parts.

Q1: Comparison of methods [20 x 3 + 10 = 70 points]

* Which method (traditional or LLM-based) does better on the three NLP tasks
* LLM

The main advantage of Logic-based Machine Learning is that because the learned knowledge is expressed as a set of logical rules, it is compact and explainable

* What issues, if any, do you see with the LLM methods

There various AI tools and each is dependent on logic. However, the logic is different for each type. Machine learning relies primarily on statistical models which uses logic to computer statistics

Expert systems use experts to determine what logic is applied. Audio visual systems use logic which can refer visual andor sound.

All people make decisions based on their perceptions which inherently is a subset of the facts. Their perception is necessarily limited to what they perceive, not necessarily reality.

At the end of the day all types of logic have advantages and disadvantages depending on the context and the most important issue is choosing how to combine them contextually.

Knowledge management is the art of turning facts into conclusions and conclusions into facts.

Q2: Based on your analysis, answer the following questions:  
[10 + 10 + 10 = 30 points]

1. Is the university safe? How did you arrive at the conclusion?

Yes:

For example

Prestory: at pre\_story1.txt

Summary :

Ongoing collaboration between agencies enables DLES to

gather data, identify crime trends on and around campus, and provides a system for notifying the Office of Student Conduct and Academic

Integrity when off-campus student crimes or policy violations occur

DLES maintains working relationships with local and state prosecutors, the court system, coroners’offices, and other investigative entities

to ensure the ongoing safety and integrity of the Carolina Community There are specific events that may occur on campus, such as the

death of a student, which would require DLES to notify the South Carolina Law Enforcement Division (SLED).

Prestory : pre\_story2.txt

Summary:

Daily Crime Log

DLES provides a daily log of all criminal reports for public inspection, which is available online or request from the DLES Records Unit

located at 1415 Henderson Street, Columbia, South Carolina 29208.The daily crime log may be found at the DLES webpage by selecting

DLES complies with federal law that prohibits the disclosure of the names of victims when publishing the daily crime log The daily crime

log includes: the nature, date, time and general location of each crime; and the disposition of the complaint, if known. If the call is placed to USC Police, Rave

Guardian automatically delivers the complete caller profile to telecommunications personnel

• Safety networks: Safety timers notifies trusted “guardians” when the user doesn’t make it to a destination and

deactivates the timer

• Anonymous crime tips: Allows dispatchers to alert the appropriate agency

page 11 Annual Security and Fire Safety Report

University of South Carolina

Division of Law Enforcement and Safety

Off-Campus:

Call If you are the victim or witness to a crime off-campus, dial 9-1-1 and the law enforcement agency responsible for that jurisdiction

will be contacted

When you contact law enforcement, you will be asked questions about what happened, including:

• Your name and telephone number;

1. Are the rights of the accuser and victim same ? If not, the policies are skewed towards whom? How did you arrive at the conclusion?

No,they are not the same. The police are skewed towards the victim.

At the summary of victim:

right treated fairness, respect, and dignity free from intimidation, harassment or abuse s constitutional rights provided by statute

reasonably informed released from custody, or has escaped

right informed and present at any criminal proceedings which are dispositive of the charges where the defendant has the right to be present

the right to be reasonably informed of and be allowed to submit either a written or oral statement at all hearings affecting bond or bail

the right to be heard at any proceeding involving a post-arrest release decision, a plea, or sentencing

protected from the accused or persons acting on his behalf throughout the criminal justice process

prompt and full restitution from the person or persons convicted of the criminal conduct that caused the victim loss or injury including both adult and juvenile offenders

the right to be informed of any proceeding any post-conviction action is being considered and be present at any postconviction hearing involving a post-conviction release decision

the right to a reasonable disposition and prompt and final conclusion of the case and

the right to have all rules governing criminal procedure and the admissibility of evidence in all criminal proceedings protect victims rights and have these rules subject to amendment or repeal by the legislature to ensure protection of these rights

At the summary of Accuser:

right to be notified in right to be assisted by various campus entities including but not limited to SAVIP, OSCAI, DLES

right to have an advisor of their choosing accompany them throughout the duration of the conduct process

right to submit a victim impact

right to have their past behaviors responsibility

right to accommodations in giving testimony consistent with providing a safe atmosphere and consistent with the rights of the accused

right to be notified in writing of the final determination and any sanction imposed on the accused as a result of the conduct process

right to receive a copy of the formal charges sent to the accused student

right to be notified of the date, time, and place of hearings

right to have the hearing authority consider as an aggravating factor when sanctioning the perpetrator provided alcohol or other drugs in the commission of a sexual assault

right to be notified of the findings and sanctions outcome of the hearing within a timeframe close to that in which the charged student was notified

1. Is it better to report a crime openly or anonymously? How did you arrive at the conclusion?

Yes,it is better.

The word count anonymous appeared more often then open.

Using Sentiment Analysis found a lot of positive words with the description of crime reported anonymously.

**Reference:**

[1] <https://github.com/huggingface/notebooks/blob/main/transformers_doc/quicktour.ipynb>