UNIVERSITY OF PETROLEUM & ENERGY STUDIES

School of Computer Science

Assignment No.1

Submitted By:

Dhruv Singhal

500075346

R177219074

AIML B3

Submitted To:

Mr. Goutam Datta

Asst. Professor

SOCS

UPES

Q1. Write a brief note on the following:

a: Deep learning for customer service:

- Deep learning in the field of eCommerce is helping hand for businesses to grow. The use of Deep Learning which is a part Machine Learning is not only helping businesses to grow but also helping customers to get a personalized feel.
- Deep Learning for customer service can be used by businesses to analyze and decrease any issue. Current DL applications that top industry leaders are employing most often include Chatbots/AI assistants: In addition, customer care bots can respond to inquiries 24 hours a day, seven days a week, without bias
- **Chatbots/AI assistants:** Natural language engagements include replying to client questions, reacting to voice instructions for simple activities, and offering product suggestions.
- **Smart organization:** Data is being analyzed using machine learning techniques to enhance inventory management.
- **Recommendation engines**: Customer activity on websites is being analyzed by companies, which are employing techniques to forecast which goods would be popular with consumers as well as provide suggestions.

b: Traditional AI vs Modern AI:

Traditional AI

Traditional AI refers to discipline of computers that show intelligent human behavior through set of algorithms and instructions. The scientists in 1956 believed that the AI technology will evolve to perform human intelligent tasks and intellect tasks as well. Prominent AI approaches back then were Expert Systems and Fuzzy Logic with Prolog and Lisp being the top choice as a programming language among C/C++. These models were lacking the computational power which let their progress to a halt or slowed down their progress. Algorithms were ready but were not able implemented due to the lack of computational power as per requirement. There were missteps and issues with AI, followed by another breakthrough by IBM when its supercomputer Deep Blue defeated world champion Garry Kasparov in New York City in 1997. Since the concept of AI was then being seen as a failure, IBM claimed it was not using AI in Deep Blue, which made for some interesting discussions.

Modern AI

The term “Data Science” was coined in early 2008. This new field of computer science has introduced advanced analytics for selected Mathematics, Opportunities, Linear Algebra and Multivariant Calculus. Later in late 2012, a real explosion occurred in Artificial Intelligence where, in the historic ImageNet competition, a CNN-based presentation called Alex Net surpassed all other competitors and received 10.8% lower errors than the previous runner-up. That was the advent of modern AI and is believed to be the cause of a new emergence in the world of AI. One major reason for the victory was the use of a Graphical Processing Unit (GPU) for neural network construction training. Later in 2015, Facebook worked hard on in-depth learning and its opportunities, along with other "AI godfathers." Today many cloud vendors are offering cloud-based GPUs for "Modern AI," and their discovery has never been an option before.

The GPU has literally changed the game by switching from the CPU to the GPU. It revolutionized technology and redefined computer power and similar processing. AI requires fast computer power due to advanced mathematical calculations. And, there has been an increasing increase in the amount of data produced over the past decade.

c: Machine Translation:

Machine Translation is a subfield of computational linguistics that includes use of software's to translate one language to another language. The challenge that MT faces is the translation of sentiment and the recognition of whole phrases and their exact or closest counterpart in the target language. The whole success of MT depends whether it can deliver the exact or closest translation possible in the target language. With evolution of technology there has been a significant growth in the Machine Translation Algorithms and techniques via extension in computational methodology. The mostly used MT approach today using Neural Machine Translation (NMT) which helps in translating and delivering the sentiments as close as possible.

Machine translation can use a method based on linguistic rules, which means that words will be translated in a linguistic way – the most suitable (orally speaking) words of the target language will replace the ones in the source language. To translate between closely related languages, the technique referred to as:

- **Rule-based machine translation**
  - Transfer-based machine translation
  - Interlingual
  - Dictionary-based
- **Statistical machine translation**
- **Hybrid Machine Translation**
- **Neural machine translation**

Q2. Discuss in detail how lexicon based and Machine Learning based approaches are useful in student's/Customer's feedback system in any academic institution/financial institution.

Student/Customer feedback when processed to find the overall gist to improve the facilities provided by the concern platform to its user is a part of sentimental analysis because feedbacks hold emotions about the facility or the service which the opted for and when it is in large quantity, there is need of a model which can input all these feedbacks and provide a dashboard or a report as in the evaluation result form which can be easily understood and contains the gist of all feedbacks at once. The success of these models depends on the that whether they are able to retain the sentiments in feedback while processing and are able to classify and point out the soft edges in the services properly or as close as possible.

Sentiment Analysis refers to finding patterns in data and inferring the emotion of the given piece of information which could be classified into one of these categories:

- Negative
- Neutral
- Positive

lexicon-based approaches for sentiment classification are based on the insight that the polarity of a piece of text can be obtained on the ground of the polarity of the words which compose it. However, due to the complexity of natural languages, a so simple approach is likely to fail since many facets of the language (e.g., the presence of the negation) are not considered. Consequently, we propose a more fine-grained approach: given a phrase **P**, we split it in several micro-phrases $m_1 \ldots m_n$ according to the splitting cues occurring in the content. As splitting cues, we used punctuations, adverbs, and conjunctions. Whenever a splitting cue is found in the text, a new micro-phrase is built

Lexicon approach includes

Opinion or review test is split into Sentences which then is processed via stemming and cleaning to form bag of words and

then are compared with Opinion Lexicon (A list of English positive and negative opinion words or sentiment words) and then passed via scoring function to provide the sentiment score.

Machine Learning approach for Sentiment Analysis,

There are few techniques and complex algorithms used to command and train machines to perform sentiment analysis. There are pros and cons to each. But, used together, they can provide exceptional results. Steps:

1. **Choose your model**
2. **Choose your classifier**
3. **Import your data**
4. **Train your classifier**
5. **Test your classifier**

A lexicon (or dictionary) represents a list of words with associated sentiment polarity. Sentiment lexicon is used to determine the polarity of a given textual content. Then, a method of computing semantic orientation of unstructured text using machine learning is used in combination to determine the sentiment of students from regular feedback cycle ranging from Positive, Neutral to Negative. The proposed method leveraged the use of sentiment lexicons which were created using semi-automatic polarity expansion algorithm. Linguistic features that have been widely used for sentiment analysis to predict sentiment from student's feedback. Naive Bayes, Maximum entropy, and Support Vector Machine (SVM) algorithms were used to train models.

Students with diversified opinions and experiences they had, when fill up the feedback forms; the lexicon and machine learning based approach has made so much simpler instead of going through every form and sorting them, this approach sorts out the overall feedback to determine the overall sentiment of the students and take constructive feedbacks into consideration to improve the system. This has also saved a lot of manual processing cost as well.

Q3. Machine Learning based techniques can be used to analyze and evaluate the credit risk data sets. Justify

Credit risk modelling is the process of estimating the probability if the party will pay back a loan. It is one of the most important mathematical problems of the financial world. The 4 basic models that can be used in credit risk modelling are:

K-Nearest Neighbor's,

Decision Trees,

Neural Networks and

Logistic Regression.

In this computation, the features such as Age, Income, consistency of employment, company, loan and default history, demography and sometimes even ethnography is determined of a customer. Then, the data is passed into the test data set obtained from years of computation of similar data and using different stated models, the credit score of a particular person is determined.

In K-nearest Neighbor's model "nearest neighbors" in the test data set is been looked upon and the credit score is determined. In decision tree, we split the dataset into smaller and smaller parts depending upon the features of the persona until we are left with the smallest dataset or a binary outcome of default or no default on the bottom of each branch.

In logistic regression, the target data is binary which is paid back or not. We find the point on the income axis that corresponds to the new person's persona, and the logistic regression line's y-value that corresponds to this point gives the probability that they will pay back the loan or not.

Neural networks are complex models which are core of the most AI models. It uses complex mathematical models to determine the credit scores of a persona.

With time and behavior in paying back loans, intervals, reasons of taking the loans (credit card companies know where the cards have been swiped), etc. Credit scores change which is solely based on the risk of paying the loans back. It determines the limit of the person, the credit card limit and the rate of interest in which he/she can get the loan all based on credit score which is risk. It is like an identity of the person in the financial institution.