**Question Bank**

**Long Questions:**

**UNIT 1**

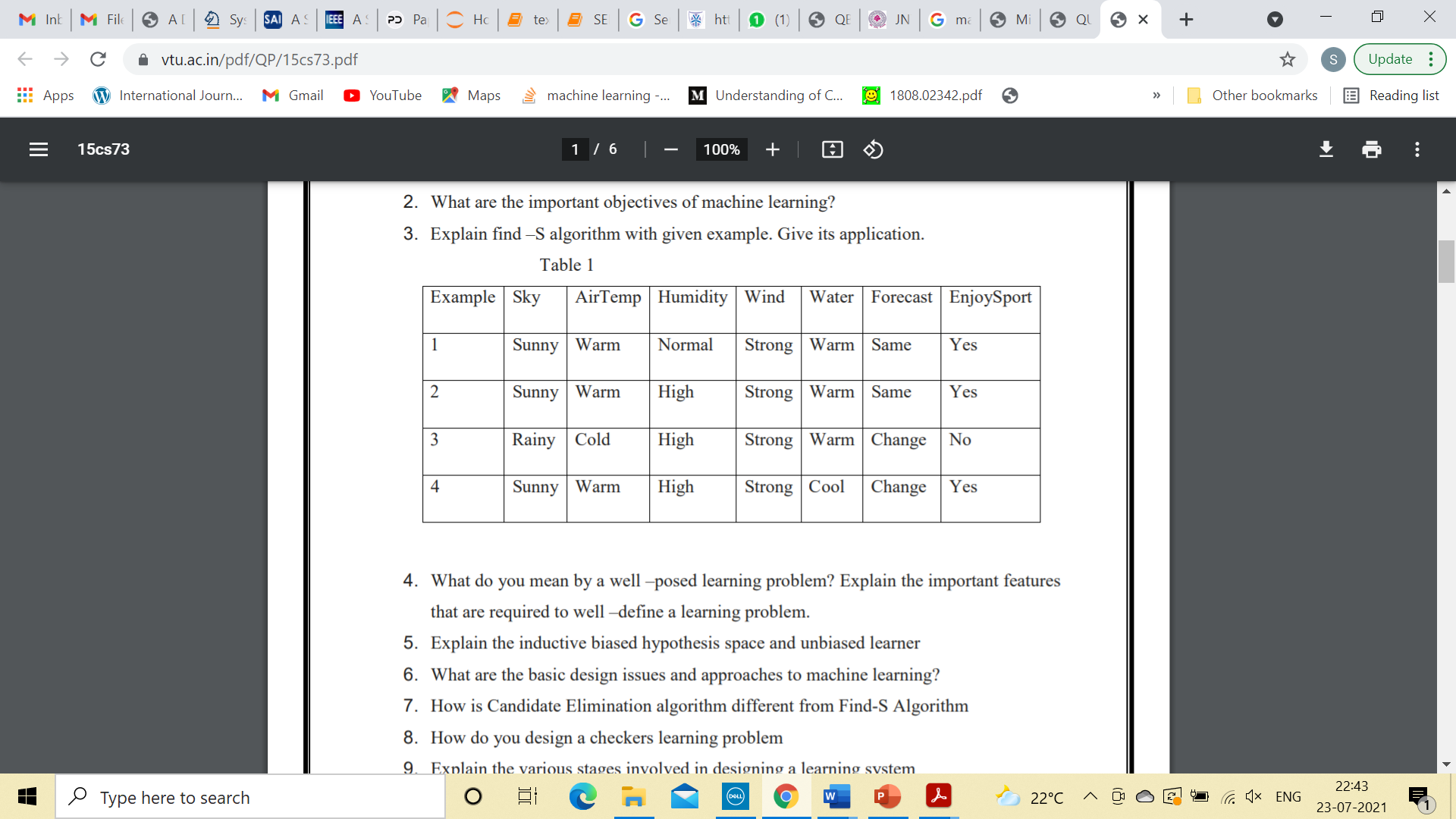
1. What are vectors in machine learning? Explain with examples the different operations performed on them.
2. Explain about matrices and the operations performed on matrices.
3. Explain Simpson’s paradox.
4. Explain about correlation and causation.
5. What are dependent and independent variables in probability?
6. Explain Bayes’ theorem.
7. Discuss about center limit theorem
8. Discuss about statistical hypothesis testing.
9. What is P-hacking
10. Discuss Bayesian Inference.

**UNIT 2**

1. Define Machine Learning. Discuss with examples why machine learning is important.
2. Discuss with examples some useful applications of machine learning.
3. What are the different types of machine learning techniques. Elaborate
4. What is train/test split of data and how it is implemented in machine learning.
5. What do you mean by regularization. Explain the different regularization methods.
6. Differentiate classification and regression
7. Explain in detail about Naïve Bayes algorithm.
8. Illustrate K-Nearest Neighbours with an example.
9. What is logistic regression?
10. Discuss about support vector machines (SVM),
11. With the help of an example explain decision trees.
12. What is meant by random forest algorithm. Explain
13. Discuss in detail about Classification Errors.

**UNIT 3**

1. Explain find –S algorithm with given example. Give its application.



1. What do u means by linear and non-liner system analysis?
2. What is a neural network explain in detail
3. Give an over view of Deep Learning.
4. Differentiate Artificial Intelligence, Machine Learning and Deep Learning.
5. What are the Characteristics of Big Data
6. What is Big Data ? Explain its Characteristics, Types, Benefits with an Example.
7. How do I extract the content from dynamic web pages?
8. What software is best suited for web scraping?
9. List out the difference between analysis and Reporting?
10. How data Science is helpful in business decision making?
11. List out the open-source tools for data science applications?

**Unit-4**

* + - 1. Write some applications of Web scraping?
      2. Explain the web scraping procedure
      3. Explain in detail about toolkits of python: NumPy, Pandas, Matplotlib, NLTK
      4. In what way the visualization of data helpful for us? Mention different visualizations and discuss about it?
      5. What is Data Munging and Explain in detail the steps of Data Cleaning
      6. Why we need Dimensionality Reduction? What are the major tools for dimensionality reduction?

**Unit 5**

1. Write a short note on Recommender Systems and distinguish the differences between the types of recommender systems?
2. Explain Collaborative recommender system with an example
3. What is the need of sentiment analysis in Text processing
4. Mention the importance of data cleaning in sentiment analysis
5. In what way, the text will be represented in order to process sentiment analysis.

**Short Questions**

**Unit 1**

1. Define Simpsons Paradox.
2. Define Correlation .
3. Write a brief note on Causation.
4. What is conditional Probability.
5. Define Hypothesis and Inference.
6. Write a short note on confidence intervals.
7. Define Bayes Theorem.
8. What is Hypothesis Testing**.**
9. Give a brief note on Random Variables.
10. What are the different operations performed on matrices.
11. What is the shape of Normalization curve?

**Unit 2**

1. Define Train/Test Split in Machine learning.
2. List the different types in Machine Learning.
3. Define Linear Regression.
4. What is Logistic Regression? Give a brief note
5. List the different regularization techniques.
6. What is the major difference between regression and classification.
7. What are classification errors?
8. Define SVM?
9. Write a short note on K-NN
10. How K-NN is different from K-Means.

**Unit-3**

1. Define Find-S algorithm.
2. What does S stand for in Find-S?
3. For What values does Find-S Algorithm performs the evaluation and why.
4. Define a neural network
5. How Machine Learning differs from Deep Learning.
6. Define Analysis.
7. Define Reporting.
8. Define Data Science and what are its applications
9. Define the traits of Big Data.

**Unit 4**

1. Define Numpy tool.
2. What is Re-Scaling? Explain the methods used for Re-scaling.
3. What is a scatter plot draw the diagram.
4. Define Data Munging
5. Define Data Cleaning.
6. List the steps for twitterAPI account creation.

**Unit 5**

1. Define sentiment analysis.
2. Define rule-based induction.
3. Define recommender systems.
4. List the different types of recommender systems.