

Student Details:

Name of the Student		Rhea Benedicta D'souza	
USN		4SO18CS097	
Semester / Section		VIII / B	
Name of the Internal Internship Guide		Ms Renuka Tantry	
Area of work		Virtual	
Internship Period	From	08/09/2021	
	To	08/10/2021	
Duration		Weeks: 6	Days: 28

Company Details:

Name of the Company	Cognitive Solution		
Address	5 th Floor, M.K. Shalimar Complex Kankanady, Mangalore - 575002		
Website	http://www.dataqueuesystems.com/cognitive/		
Company Head	Mr. Shameer Ahmed		
Name of the Industry Guide	Ms. Siby Sussan		
Contact No	+91 9482579584		
Email - ID	Sibysusan.abr@gmail.com		

Ms.Siby Sussan

Name & Signature of the Internal Guide

Name of the External Guide

**Name & Signature of the Internship
Coordinator**

Name & Signature of the HOD

VISION OF THE DEPARTMENT

To be recognized as a centre of excellence in computer and allied areas with quality learning and research environment.

MISSION OF THE DEPARTMENT

1. Prepare competent professionals in the field of computer and allied fields enriched with ethical values.
2. Contribute to the Socio-economic development of the country by imparting quality education in computer and Information Technology.
3. Enhance employability through skill development.

Undergraduate Programme in Computer Science and Engineering (B.E.)

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

- I. To impart to students a sound foundation and ability to apply engineering fundamentals, mathematics, science and humanities necessary to formulate, analyze, design and implement engineering problems in the field of computer science.
- II. To develop in students the knowledge of fundamentals of computer science and engineering to work in various related fields such as network, data, web and system engineering.
- III. To develop in students the ability to work as a part of team through effective communication on multidisciplinary projects.
- IV. To train students to have successful careers in computer and information technology industry that meets the needs of society enriched with professional ethics.
- V. To develop in students the ability to pursue higher education and engage in research through continuous learning.

PROGRAMME OUTCOMES (POs)

By the end of the undergraduate programme in CSE, graduates will be able to:

1. **Engineering Knowledge:** Apply knowledge of mathematics, science, engineering fundamentals, computer science and engineering to solve complex engineering problems.
2. **Problem Analysis:** Identify, formulate, research literature, and analyze complex engineering problems in reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/Development of Solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct Investigations:** Conduct investigations of complex problems using research based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **Engineer and Society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and Sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and Teamwork:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, give and receive clear instructions.
11. **Project Management and Finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a

member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Lifelong Learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAMME SPECIFIC OUTCOMES (PSOs)

By the end of the undergraduate programme in CSE, graduates will be able to:

1. **Entrepreneurship and Freelancing:** Understand the principles underlying entrepreneurship, freelancing and the requirements to initiate a start up in the IT or related domains.
2. **Competitive Exams:** Participate effectively in competitive examinations related to certification, career growth and admission to higher studies.

COURSE OUTCOMES (COs)

18CSI85.1	Explain the industry practices and the tools used with respect to the project carried out.
18CSI85.2	Demonstrate problem-solving skills gained through the internship experience.
18CSI85.3	Demonstrate effective verbal and written communication skills.
18CSI85.4	Collaborate with teams as a team leader/ team member during the course of the project and build a professional network.
18CSI85.5	Illustrate the various ways of effective management of personal behavior, ethics and attitude.
18CSI85.6	Illustrate the various ways of adaptation to changing needs while working in the industry.

CO – PO Mapping

Keywords (PO/PSO)	Apply Knowledge	Solve Problems	Design/ Development of Solution	Conduct Investigations	Use Modern Tools	Engineer and Society	Environment and Sustainability	Professional Ethics	Individual and Team Work	Communicate Effectively	Project Management and Finance	Lifelong Learning	Entrepreneurship, Freelancing & start-up	Competitive Exams & Higher Studies
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
18CSI85.1	-	-	-	-	3	-	-	-	-	-	3	-	3	-
18CSI85.2	-	-	-	-	-	-	-	-	-	2	3	-	3	-
18CSI85.3	-	-	-	-	-	-	-	-	-	3	2	-	2	-
18CSI85.4	-	-	-	-	-	-	-	2	-	-	3	-	3	-
18CSI85.5	-	-	-	-	-	-	-	3	-	-	1	-	2	-
18CSI85.6	-	-	-	-	2	-	-	-	-	-	2	-	2	-

INTERNSHIP WORK PLAN

Area of Work	Virtual
Internship Topic	Machine learning using python
Objectives of the Internship	The objective was to explore the differences between supervised and unsupervised learning techniques
Real Time Applications	NIL
Expected Outcomes	<p>The expected outcomes were to learn Machine Learning algorithms and apply them such as:</p> <ul style="list-style-type: none">• KNN algorithm• KMeans algorithm• Logistic regression• Decision Tree classification
Skills acquired during Internship	<p>Technical</p> <ul style="list-style-type: none">• Pre-process a dataset and prepare it for machine learning algorithms• Classification, linear regression and logistic regression• Clustering and decision tree <p>Non-technical</p> <ul style="list-style-type: none">• Communication skills• Task management
Challenges faced during Internship	No challenges faced
Any other Comments	No comments

Daily Work Plan

Week 1: Understanding Python

Date	08/09/2021 - 11/09/2021
Task Assigned	Basics of Python
Task Objective	To understand the basics of python
Task Outcome	To be able to comprehend the basics of python
Brief Description of the Work	
In week 1, in order to move ahead to machine learning with AI, basic python concepts had to be brushed up. I had to install the python software with Anaconda and set up the environment variables. This was followed by the facilitator teaching us the syntax, data types, operators and different data structures in python. I was also given few exercises based on topics taught.	

Week 2: Understanding Python

Date	13/09/2021 - 16/09/2021
Task Assigned	Basics of Python
Task Objective	To understand the basics of python
Task Outcome	To be able to comprehend the basics of python
Brief Description of the Work	
In week 2, after the basic concepts were introduced, I was taught concepts such as loops, object oriented programming, JSON, and Regular Expressions. I was given exercises on these concepts. The facilitator then introduced us to the basics of NumPy, Pandas and Matplotlib.	

Week 3: Introduction to Machine Learning

Date	20/09/2021 - 23/09/2021
Task Assigned	Understanding machine learning concepts
Task Objective	To understand the concepts of machine learning
Task Outcome	To be able to comprehend and apply the concepts of machine learning

Brief Description of the Work

In week 3, I was introduced to machine learning. This included the difference between Artificial Intelligence and machine learning, how to get datasets, importing libraries, the need for data preprocessing, the different types of machine learning namely: supervised and unsupervised, and the difference between these two types of machine learning techniques. Supervised Learning was further explained by introducing us to concepts like regression analysis: its terminologies, use, and types of regression.

I was then taught about linear regression, how it makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc and then about simple linear regression and multiple linear regression. I was then given a demo of implementation of Simple Linear regression. A dataset consisting of two variables was taken: salary (dependent variable) and experience(independent variable).

The goals of the problem was as follows:

- We want to find out if there is any correlation between these two variables
- We will find the best fit line for the dataset.
- How the dependent variable is changing by changing the independent variable.

This was implemented using a simple linear regression model to find out the best fitting line for representing the relationship between these two variables.

The solution steps were as follows:

Step-1: Data Pre-processing

Step-2: Fitting the Simple Linear Regression to the Training Set

Step-3. Prediction of test set result

Step-4. visualizing the Training set results

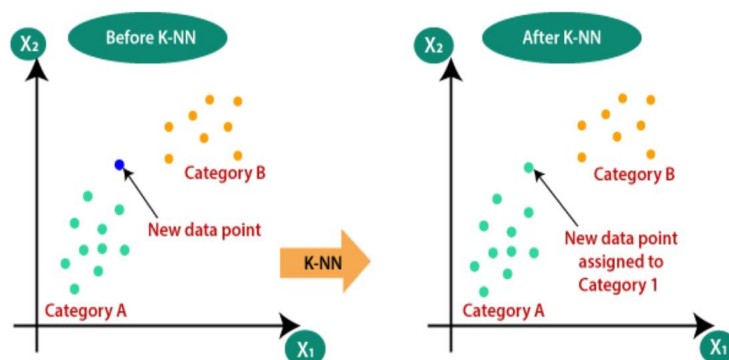
Week 4: K-NN algorithm, SVM algorithm, naive Bayes Classifier

Date	27/09/2021 - 30/09/2021
Task Assigned	Basics of Python
Task Objective	To understand the, K-NN algorithm, SVM algorithm, Naive Bayes Classifier concepts
Task Outcome	To be able to comprehend and apply the K-NN algorithm, SVM algorithm, Naive Bayes Classifier concepts

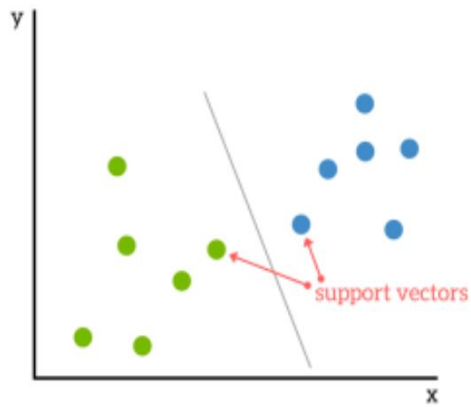
Brief Description of the Work

In week 4, I was introduced to the classification algorithms. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, Yes or No, 0 or 1, Spam or Not Spam, cat or dog, etc. Classes can be called as targets/labels or categories.

K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. Brief example: Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x_1 , so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset.



A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. I was taught about keywords called 'hyperplane' which is a line that linearly separates and classifies a set of data and 'Support vectors' which are the data points nearest to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data set.. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes. In the below figure, the grey line is the hyperplane.



Naive Bayes is a probabilistic machine learning algorithm based on the Bayes Theorem, used in a wide variety of classification tasks. Example: Suppose we have a dataset of weather conditions and corresponding target variable "Play". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions.

	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

Applying Bayes'theorem:

$$P(\text{Yes}|\text{Sunny}) = P(\text{Sunny}|\text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$$

$$P(\text{Sunny}|\text{Yes}) = 3/10 = 0.3$$

$$P(\text{Sunny}) = 0.35$$

$$P(\text{Yes}) = 0.71$$

$$\text{So } P(\text{Yes}|\text{Sunny}) = 0.3 * 0.71 / 0.35 = \mathbf{0.60}$$

$$P(\text{No}|\text{Sunny}) = P(\text{Sunny}|\text{No}) * P(\text{No}) / P(\text{Sunny})$$

$$P(\text{Sunny}|\text{No}) = 2/4 = 0.5$$

$$P(\text{No}) = 0.29$$

$$P(\text{Sunny}) = 0.35$$

$$\text{So } P(\text{No}|\text{Sunny}) = 0.5 * 0.29 / 0.35 = \mathbf{0.41}$$

So as we can see from the above calculation that $P(\text{Yes}|\text{Sunny}) > P(\text{No}|\text{Sunny})$

Hence on a Sunny day, Player can play the game.

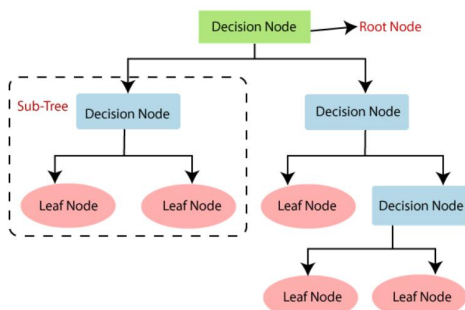
Week 5: Logistic Regression, Decision Tree classification, Kmeans clustering

Date	01/10/2021 - 04/10/2021
Task Assigned	Logistic Regression, Decision Tree classification, Kmeans clustering
Task Objective	To understand the Logistic Regression, Decision Tree classification and Kmeans clustering concepts
Task Outcome	To be able to comprehend and apply Logistic Regression, Decision Tree classification and Kmeans clustering concepts

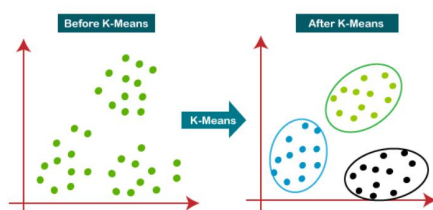
Brief Description of the Work

In week 5, I was taught about logistic regression, decision tree classification and kmeans clustering. I was taught about the difference between logistic and linear regression. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.



Kmeans is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties.



The trainer showed some examples of the applications of kmeans clustering for my better understanding.

As the internship duration was coming to an end, I was asked to choose any algorithm of my

interest and complete a mini project on a topic I desired. I was informed that the presentation would take place on the 8th of October, 2022. The trainer informed me of the task on the 4th of October.

Week 6: Project work

Date	05/10/2021 - 08/10/2021
Task Assigned	Completion of project
Task Objective	To understand the concepts taught during internship and apply them in the project
Task Outcome	To be able to apply the concepts taught and apply them in project

Brief Description of the Work

In week 6, I went through some of the topics taught and chose my topic of interest and started working on spam mail detection. This project used a logistic regression model to train the mail dataset. I created a kaggle account and chose one of the dataset which contained a good number of spam and ham mails. I altered the dataset by taking some mails from my gmail account and adding them onto the dataset. The next two days I spent on my project, by performing the following steps: importing dependencies, data collection and pre processing, label encoding, splitting data into training and testing, feature extraction, training the model, evaluating the trained model and building a predictive system. I successfully completed my project on time by the 7th of October 2022.

Internship Closure Report

<i>A brief description of the internship outcomes achieved</i>	
Internship Objectives:	The objective of this internship was to learn machine learning & AI techniques using python. Also to learn on how to explore the differences between supervised and unsupervised learning techniques, introducing Machine Learning algorithms and then apply it to the project work.
Objectives Accomplished:	Objectives accomplished were gaining knowledge of algorithms, understanding it and then applying it to the project. Also I learnt supervised and unsupervised techniques taught in this internship.
Objectives could not be Accomplished:	All the objectives were accomplished
Reasons for Non-Accomplishment	No Reasons
Sills acquired during internship period	Technical <ul style="list-style-type: none">• Pre-process a dataset and prepare it for machine learning algorithms• Classification, linear regression and logistic regression• Clustering and decision tree Non-technical <ul style="list-style-type: none">• Communication skills• Task management

Challenges faced during internship Period	No challenges faced
Overall Outcome of Internship Training	The overall outcome of this internship was to learn about the machine learning algorithms and techniques and apply them in a project.

Signature of the student with Date

FACULTY INCHARGE REMARKS

About the Company:

About Student Performance:

Signature with Date