ST.JOSEPH’S COLLEGE FOR WOMEN (AUTONOMOUS), VISAKHAPATNAM

V SEMESTER **COMPUTER SCIENCE** Time: 3Hrs/Week

CS-E3-5603(3) **DATA SCIENCE**  Max.Marks:100

w.e.f. 20AHBatch **SYLLABUS**

**COURSE OBJECTIVES: -**

* To learn the basics of data science
* Implement it in real-world environment
* Learn basic programming using data science.

**COURSE OUTCOMES:**

* Students after successful completion of the course will be able to:
* Develop programming abilities.
* Demonstrate statistical analysis of data.
* Develop the ability to build data-based models.
* Improve their skill in data management.
* Apply data science concepts and methods to solve problems in real-world contexts.

**UNIT - I: Introduction:** The Ascendance of Data, What is Data Science? , Finding key Connectors, Data Scientists You May Know, Salaries and Experience, Paid Accounts, Topics of Interest, Onward.

**Python:** Getting Python, The Zen of Python, Whitespace Formatting, Modules, Arithmetic, Functions, Strings, Exceptions, Lists, Tuples, Dictionaries, Sets, Control Flow, Truthiness, Sorting, List Comprehensions, Generators and Iterators, Randomness, Object – Orienting Programming, Functional Tools, enumerate, zip and Argument Unpacking, args and kwargs, Welcome to Data Sciencester!

**Visualizing Data:** matplotlib, Bar charts, Line charts, Scatterplots.

**Linear Algebra:** Vectors, Matrices

**UNIT -II: Statistics:** Describing a Single Set of Data, Correlation, Simpson’s Paradox, some Other Correlation Caveats, Correlation and Causation.

**Probability:** Dependence and Independence, Conditional Probability, Bayes’s Theorem, Random Variables, Continuous Distributions, The Normal Distribution, The Central Limit Theorem.

**Hypothesis and Inference:** Statistical Hypothesis Testing, Example: Flipping a Coin, Confidence Intervals, P-hacking, Example: Running an A/B Test, Bayesian Inference.

**Gradient Descent:** The Idea behind Gradient Descent, Estimating the Gradient, Using the Gradient, Choosing the Right Step Size, Putting It All Together, Stochastic Gradient Descent.

**UNIT -III :Getting Data:** stdin and stdout, Reading Files – The Basics of Text Files, Delimited Files, Scraping the Web - HTML and the parsing Thereof, Example: O’Reilly Books About Data, Using APIs – JSON ( and XML), Using an Unauthenticated API, Finding APIs.

**Working with Data**: Exploring Your Data, Exploring One-Dimensional Data, Two Dimensions Many Dimensions, Cleaning and Munging, Manipulating Data, Rescaling, Dimensionality Reduction.

**Machine Learning**: Modeling, What Is Machine Learning? Over fitting and under fitting, Correctness, The Bias-Variance Trade-off, Feature Extraction and selection.

**P.T.O.**

CS-E3-5603(3) ::2::

**UNIT-IV: K-Nearest Neighbors:** The Model, Example: Favorite Languages, The Curse of Dimensionality.

**Naive Bayes**: A Really Dumb Spam Filter, A More Sophisticated Spam Filter, Implementation, Testing Our Model.

**Simple Linear Regression:** The Model, Using Gradient Descent, Maximum Likelihood Estimation.

**Multiple Regression:** The Model, Further Assumptions of the Least Squares Model, Fitting the Model, Interpreting the Model, Goodness of Fit.

**UNIT–V: Logistic Regression:** The Problem, The Logistic Function, Applying the Model, Goodness of Fit Support Vector Machines.

**Decision Trees:** What Is a Decision Tree? Entropy, The Entropy of a Partition, Creating a Decision Tree, Putting It All Together, Random Forests.

**Neural Networks:** Perceptron, Feed-Forward Neural Networks And Back propagation, Example: Defeating a CAPTCHA.

**Clustering:** The Idea, The Model, Example: Meetups , Choosing k, Example: Clustering Colors, Bottom-up Hierarchical Clustering.

**REFERENCES:**

1. Data Science from Scratch by Joel Grus O’Reilly Media .
2. Wes McKinney, “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython”, O’Reilly, 2nd Edition, 2018.
3. Jake VanderPlas, “Python Data Science Handbook: Essential Tools for Working with Data”, O’Reilly, 2017.
4. Web resources:
   1. https://www.edx.org/course/analyzing-data-with-python

b.http://math.ecnu.edu.cn/~lfzhou/seminar/[Joel\_Grus]\_Data\_Science\_from\_Scratch\_First\_Princ.pdf

1. Other web sources suggested by the teacher concerned and the college librarian including reading material.

\*\* \*\* \*\*