ST. JOSEPH'S COLLEGE (AUTONOMOUS)

IRINJALAKUDA

CURRICULA AND SYLLABI FOR

`

Master of Data Analytics

Under Choice Based Credit &Semester System

2021 Admission

St. Joseph's College (Autonomous), Irinjalakuda

Department of Computer Science

Board of Studies in Computer Science

SL.NO	NAME	OFFICIAL ADDRESS	DESIGNATION
1.	Siji P D	Assistant Professor ,St. Joseph‟s College, Irinjalakuda,Kerala 680121	Chairman
2.	Dr. Soniya Sunny	Associate Professor ,PrajyothiNikethan College Pudukadu	University Nominee
3.	Dr.Vinod Chandra S. S	Associate Professor &Director,Computer Centre, University of Kerala, Thiruvananthapuram – 695034	Subject Expert
4.	Dr.CiniKurian	Associate Professor ,Al-ameen College, Edathala, AluvaErnakulam (Dist.), Kerala - 683 564	Subject Expert
5.	Mr. Sujith. E. S	Senior Associate of Projects, Cognizant Technology Solutions(CTS),Infopark, Kochi	Industry Expert
6.	Ms. Sreelakshmi P	Asst. Manager /Cyber Risk-Risk Advisory,DeloitteTouche,Tohmatsu India LLP,9th Floor, Edapilly	Alumnus
7.	Ms. Reesha P U	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
8.	Ms. Ambily Jacob	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
9.	Ms. Magniya Davis	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
10.	Ms. Jasmine Jose	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
11.	Ms. PrameejaPrasidhan	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member

12.	MsGeethu Wilson	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
13.	Ms .Anu Jose	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
SL.NO	**NAME**	**OFFICIAL ADDRESS**	**DESIGNATION**
1.	Siji P D	Assistant Professor ,St. Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
2.	Dr. Soniya Sunny	Associate Professor ,PrajyothiNikethan College Pudukadu	Subject Expert
3.	Dr.Vinod Chandra S. S	Associate Professor &Director,Computer Centre, University of Kerala, Thiruvananthapuram – 695034	Subject Expert
4.	Dr.CiniKurian	Associate Professor ,Al-ameen College, Edathala, AluvaErnakulam (Dist.), Kerala - 683 564	Subject Expert
5.	Mr. Sujith. E. S	Senior Associate of Projects, Cognizant Technology Solutions(CTS),Infopark, Kochi	Industry Expert
6.	Ms. Sreelakshmi P	Asst. Manager /Cyber Risk-Risk Advisory,DeloitteTouche,Tohmatsu India LLP,9th Floor, Edapilly	Alumnus
7.	Ms. Reesha P U	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
8.	Ms. Ambily Jacob	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
9.	Ms. Magniya Davis	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
10.	Ms. Jasmine Jose	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
11.	Ms. PrameejaPrasidhan	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
12.	Ms. Geethu Wilson	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
13.	Ms .Anu Jose	Assistant Professor, St.Joseph‟s College, Irinjalakuda,Kerala 680121	Faculty member
14.	Ms. Anaida Simon	Association Secretary,Department of Computer Science,St.Joseph‟s College, Irinjalakuda,Kerala 680121	Student Representativ e

Contributors towards Curriculum and Syllabus

FOREWORD

Higher Education scenario in Kerala has been going through turbulent transformations in recent times with the grant of autonomy to colleges by the State Government. There is no doubt about the qualitative worth of the institutions handpicked for autonomy. However, there are apprehensions about the absorption and implementation of the package of autonomy. St.Joseph‟s College was given

autonomy in the year 2016, and has since then been endeavoring to reinvent itself.

Academic autonomy has given us the freedom to recreate our own curriculum and syllabus keeping in mind the challenges and changing needs of the society, the nation, the industry and the world. Hence, a structured feedback on the requirements of the new millennium was sought from all the relevant stakeholders of the institution- students, faculty, alumnae, parents, industry experts, employers etc.

The suggestions of the stakeholders were incorporated into the curricula and syllabi, and presented in the respective Boards of Studies for discussion. The changes pointed out were duly considered and the restructured syllabi are then presented to, and ratified by, the Academic Council.

The role of the IQAC of the college in the above exercise is laudatory. The Cell spearheads all the quality enhancement endeavours, including that of curriculum and syllabus redesigning. By organizing workshops, seminars and hands on training sessions, the cell has facilitated a smooth conduct of the restructuring process. At the end of the year, an evaluation of the syllabi followed is also undertaken, with suggestions noted down for future changes.

As an institution that wishes seriously to provide enhanced quality education to young women students in order to empower them to be fit for the changing world, St.Joseph‟s College is bravely facing the challenges even as it is happily handling the possibilities, that autonomy has brought to it. Academic enriching programmes, skill – based micro credentials, ICT up gradations, promotional activities for a culture of research, etc are a few of the multifarious responsibilities invested with the college in its restructuring of curriculam and redesigning of syllabus.

I specially thank the IQAC, the Heads of various departments the faculty, and staff, directly in charge of the syllabus updation, for their sincere and dedicated efforts.

Principal

ACKNOWLEDGEMENT

The syllabus restructuring of the M.SC in Data Analytics programme would not have been possible without the guidance and the help of several individuals who in one way or other contributed and

extended their valuable assistance in the preparation and completion of this work. The Board of Studies in Department of Computer Science takes this opportunity to express our deep appreciation to all academicians and professionals who participate in the workshops organized by St. Joseph‟s College (Autonomous) for restructuring the PG Course in Department of Computer Science. I remember with gratitude the support of our Principal, Dr. Sr.Lissy Anto P, Dr.Saritha Namboothiri, Senate member in the University of Calicut, Associate Professor and Head, Department of computer science, Sreekrishnapuram V.T.Bahatathirippad College. I am grateful to the Board of Studies members of the Department of Computer Science for their valuable insights and guidance throughout the process. I am indebted to the faculties of Department of Computer Science for their kind co-operation in all phases of this syllabus restructuring process. We place on record our gratitude to the IQAC Co-ordinator Dr.Naijil George, Assistant Professor of Department of Biotechnology, for the timely and valuable guidance. We express our whole- hearted gratitude to all those who have helped us in this endeavor.

CONTENT

ST. JOSEPH'S COLLEGE, (AUTONOMOUS), IRINJALAKUDA

DEPARTMENT OF COMPUTER SCIENCE

2020 ADMISSION

Preface

MSC.Data Analytics Programme at Department of Computer Science, St.Joseph"s College(Autonomous), Irinjalakuda was following the syllabus of University of Calicut, from the academic year 2013-2014. After it was granted academic autonomy in the year 2015, and hence has the privilege of restructuring the syllabus. Keeping an eye on the industry and to modernize the curriculum, the Board of Studies members of the Department of Computer Science, St.Joseph's College, (Autonomous), Irinjalakuda has initiated restructuring of the syllabus for MSc.Data Analytics programme for 2020 Admissions. The syllabus aims to focus on enabling the students to familiarize with the new technologies, and at the same time enhance and strengthen the fundamental knowledge in Computer Science and Mathematics.

STUDENT ATTRIBUTES

The motto of the institution is "Light, Life, Love"
Light for the illumination of the heart and mind
Life for the fullness of growth – physical, mental, intellectual and spiritual
Love for fellowship with the Supreme & with one another

The motto enshrines the vision of the Founders for the students and constitutes the foundation for the acquisition of the following student attributes envisioned by the institution.

Empowerment
Life Long Learning
Holistic Development
Value Orientation

Social Responsibility
Nation Building Capacity
Green Thinking
Creativity & Innovation
Acquiring Life Skills

- Discipline
- Leadership / Team skills
- Problem solving skills
- Communicability

The above Student Attributes will be attained in the span of their student life at St. Joseph‟s College through various activities such as

- Curricular, Co-curricular & extra-curricular
- Sports, games, fine arts and cultural
- Enrichment / certificate courses
- Extension / outreach programmes
- Healthy / Best practices

PROGRAMME OUTCOMES

At the end of a post graduate programme, the student would have :

1. Acquired the ability for critical thinking and problem solving
2. Attained life skills and communication skills
3. Inculcated moral and ethical values
4. Become a promoter of unpolluted environs and proactive society
5. Developed a culture of research and lifelong learning
6. Become an empowered woman aware of global perspectives and national realities

PROGRAMME SPECIFIC OUTCOME

	Program Specific Outcomes
PSO1	Learn Data analytics tools and techniques, SQL databases, the languages of R and Python, data visualizations, statistics and predictive analytics in a business environment.
PSO2	Explore basic programming skills and how to wrangle data from diverse sources and shape it to enable data-driven applications

PSO3	Ability to preprocess and analyse various data, latest trends in technology development and thereby innovate new ideas and solutions to existing problems.
PSO4	Data analytics is the science of analyzing raw data in order to make conclusions about that information. Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information
PSO5	Ability to apply mathematical / statistical implementation, manage Projects / Internship and function effectively as an individual, and as a member or leader in diverse teams
PSO6	Apply knowledge of mathematics, statistics, science and computing/programming skill to model the software applications, configure software platform and analyze real time data in a heterogeneous domain.

AIMS AND OBJECTIVES

First Semester

- To understand Data source evolution, data Characteristics and data processing models.
- To understand and apply data process architecture HADOOP, SPARK and write MapReduce programs.
- To analyze Big Data use cases for specific domain and applications.
- To understand the Probability Theory
- To understand regression and theoretical distributions

- To understand the concepts of Data Warehouse architecture and apply for various domains.
- To understand Data Mining techniques Cluster, Classification and Association Rule Mining.
- To understand the concepts of Web mining, Text mining and Spatial mining.
- To understand the basics of Python Data structures and Programming constructs.
- To understand and Apply Python Libraries for Data Science and Machine Learning
- To understand the object-oriented concepts: Class, Inheritance and Polymorphism.
- To understand and analysis concepts of Algorithmic analysis and algorithm approaches

Second Semester

- To understand the concepts of DBMS, Data Model and Normal forms**.**
- To understand the concepts of concurrency control and Recovery.
- To understand basics of SQL and NoSQL databases.
- To understand and apply MongoDB (NoSQL) for Data Analysis using CURD and User Management.
- To understand linear programming methods
- To understand Dynamic programming approach
- To understand the basics constructs of R Programming and Visualization.
- To understand and apply Exploring variables using Visualization.
- To understand and apply Inferential Statistics and Regression Models
- To understand how accurately represent voluminous complex data set in web and from other data sources
- To understand the methodologies used to visualize large data sets
- To know how to work with visualization tools.
- To understand concepts basics concepts of Linear Algebra
- To understand concepts of vector spaces and matrices
- To understand the applications of Linear Algebra in Machine Learning

Third Semester

- To understand cloud service models, architecture, programming model, security and familiarize with leading service cloud providers to critically evaluate the cloud services for business applications of various domains
- To understand MapReduce programming architecture, processing models.
- To understand and design MapReduce Programming using PIG and Hive
- To understand and compare the architectural and processing of MapReduce Programing languages Pig, Hive and SPARK
- To understand the concepts of machine learning concepts, algorithms, and probabilistic, linear, graphical models.

- To apply the machine learning algorithms for various applications.

Fourth Semester

- Internship or Project to get hands on experience on what the students has learned.
- To apply, analyze and understand to implement the various techniques that have been studied throughout the Data Analytics course

COURSE DESIGN

The MSC.Data Analytics programme includes

i. Core courses
ii. Elective Courses
iii. Audit courses

The number of Courses for the MSC. Data Analytics programme contains 16 compulsory core courses and 4 elective courses from the frontier area of the core courses, 2 audit courses and a project. Project Work / Internship are mandatory for MSC Data Analytics programme and these shall be done in the end of fourth semester.

Duration of the programme

The minimum duration for completion of a four semester PG Programme is two years.
The maximum period for completion is 4 years. The duration of each semester shall be 90 working days, inclusive of examinations, spread over five months. Odd semesters shall be held from June to October and even semesters from November to March subject the academic calendar of the St.Joseph‟s College(Autonomous), Irinjalakuda

Programme Structure

The MSc. Data Analytics programme includes three types of courses, viz., Core courses (Code C), Elective Courses (Code E) and Audit Courses (Code A). Every student of the MSc Data Analytics programme shall have to work on a project/dissertation of not less than 8 credits under the supervision of a faculty member as per the curriculum. Project/dissertation shall be treated as Core Courses. Project Work is mandatory for all regular programmes and Comprehensive Viva-voce is optional and these shall be done in the end semester. The combined Credit for the Project Work and Comprehensive Viva-voce shall not be more than 8 (eight) credits subject to a minimum of 4 (four) credit for Project Work. All students have to submit a Project Report/Dissertation in the prescribed structure and format as a part of the Project Work undertaken .Total credit for the programme shall be 80 (eighty), this describes the weightage of the course concerned and the pattern of distribution is as detailed below

Programme Duration	4 Semester

Core Courses	16
Elective Courses	4
Project/dissertation	1
Comprehensive Viva-voce	
Minimum Attendance required	75%

Elective courses shall be spread over either in the Third & Fourth Semesters combined. Audit Courses: There will be two Audit Courses (Ability Enhancement Course & Professional Competency Course) with 4 credits each. These have to be done one each in the first two semesters. The credits will not be counted for evaluating the overall SGPA & CGPA. The colleges shall conduct examination for these courses and have to intimate /upload the results of the same to the University on the stipulated date during the III Semester. Students have to obtain only minimum pass requirements in the Audit Courses.

Study Tour / Field visit / Industrial visit / Trip for specimen collection may be conducted as a part of the Programme.

Semester	Course Title	Suggested Area
I	Ability Enhancement Course (AEC)	Internship / Seminar presentation / Publications / Case study analysis / Industrial or Practical Training/Community linkage programme /Book reviews etc.
II	Professional Competency Course (PCC)	To test the skill level of students like testing the application level of different softwares such as SPSS/R/ Econometrics / Python/Any software relevant to the programme of study / Translations etc.

Courses and Credit distribution

The required number of credits as specified in the syllabus/regulations must be acquired by the student to qualify for the degree. A student shall accumulate a minimum of 80 credits for the successful completion of the MSc. Data Analytics programmes.

Semester	Course	Teaching Hours	Credit
I	Core Courses (Theory/Practical)	*715+12=727*	*(52+6+8)=66*
II	Elective Courses (Theory/Practical)	*18*	*14*
III	Including: ● Comprehensive Viva-voce (Optional) ● Project Work / Dissertation	*15*	*8*
Total credit			

Audit Courses

In addition to the above courses there will be two Audit Courses *(Ability Enhancement Course & Professional Competency Course)* with 4 credits each. The college will conduct examinations for these courses in respective semesters and intimate /upload the results of the same to the Controller of Examinations of St. Joseph"s College (Autonomous) Irinjalakuda. The College will intimate/upload the results of the same to the University on the stipulated date during the third semester. The credits will not be counted for evaluating the overall SGPA & CGPA. The details of Audit courses are given below.

Semester	Course	Teaching Hours	Credit
I	***Audit Course I :*** *Ability Enhancement Course (AEC)*	*0*	*4*
II	***Audit Course II :*** *Professional Competency Course (PCC)*	*0*	*4*

Project Work / Dissertation & Comprehensive Viva-Voce

There is a Project work with dissertation and Comprehensive Viva-Voce as separate courses relating to the core area under study in the end Semester and included in the Core Courses. Viva-voce related to Project work is one of the criteria for Project Work evaluation. Students have to submit a Project Report / Dissertation in the prescribed structure and format as a part of the Project Work undertaken. There will be External and Internal evaluation for Project Work/ Comprehensive Viva-Voce and these shall be combined in the proportion of 4:1.

COURSE CODE FORMAT

The following are the common guidelines for coding various courses in order to get a uniform identification. It is advisable to assign a nine Digit Code (combination of Alpha Numerical) for various courses as detailed below:

1. **First two digits** indicate the code of college SJ
2. **Next three digits** indicate the Programme/discipline code (ENG for English, MCM for M.Com, CHE for chemistry, PHY for physics, MLM for Malayalam, SKT for Sanskrit, HTY for History etc.)
3. **Sixth digit** is the Semester indicator which can be given as 1, 2, 3 & 4 respectively for I, II, III & IV Semester (MCM1, CHE2 Etc).
4. **Seventh digit** will be the Course Category indicator as detailed below :

Sl No	**Nature of Course**	**ourse Code**
1	**C**ore Courses	**C**
2	**E**lective Courses	**E**
3	**P**roject	**P**
4	Comprehensive **V**iva	**V**
5	Practical / **L**ab	**L**
6	**A**udit Courses	**A**

5. **Last two digits** indicate the serial number of the respective courses. If there is one digit it should be prefixed by '0'(Zero). (01, 02, etc)
6. If the number of courses in one category is only one (eg : Viva, Project etc.), assign the course serial

number as 01.

7. Examples :

Sl. No	Code	Details
1	SJMCM 1C01	M.Com I Sem Core Course No1
2	SJCHE 2 A 02	Chemistry II Sem Audit Course No.2
3	SJENG 4 V01	English IV Sem Viva No. 1
4	SJMLM 3 E02	Malayalam III Sem Elective No. 2
5	SJPHY 4 P 01	Physics IV Sem Project Work No. 1
6	SJ BGY 2 L 02	Biology II Sem Practical No. 2
7	SJPSY 3 C 02	Psychology III Sem Core Coure No. 2
8	SJHTR 2 E 01	History II Sem Elective Course No. 1

EVALUATION AND GRADING

The evaluation scheme for each course shall contain two parts; (a) Internal / Continuous Assessment (CA) and (b) External / End Semester Evaluation (ESE). Of the total, 20% weightage shall be given to Internal evaluation / Continuous assessment and the remaining 80% to External/ESE and the ratio and weightage between Internal and External is 1:4.

Primary evaluation for Internal and External shall be based on 6 letter grades (A+, A, B, C, D and E) with numerical values (Grade Points) of 5, 4, 3, 2, 1 & 0 respectively. Grade Point Average: Internal and External components are separately graded and the combined grade point with weightage 1 for Internal and 4 for external shall be applied to calculate the Grade Point Average (GPA) of each course. Letter grade shall be assigned to each course based on the categorization based on Ten point Scale.

Evaluation of Audit Courses

The examination and evaluation shall be conducted by the college itself either in the normal structure or MCQ model from the Question Bank and other guidelines provided by the University/BoS. The Question paper shall be for minimum 20 weightage and a minimum of 2 hour duration for the examination. The marks of audit courses one and two will be forwarded to Controller of Examinations of St. Joseph‟s College (Autonomous) Irinjalakuda in time of respective semesters. The result has to be intimated / uploaded to the University during the Third Semester as per the notification of the University.

Phases for Evaluation:

***I Phase**: To be done by the concerned Teacher/Examiner based on 6 Point Scale*

1. Evaluation of all individual External Theory courses and Internal evaluation
2. Evaluation of Project Work External and Internal
3. Evaluation of External and Internal Practical Courses
4. Evaluation of External and Internal Comprehensive Viva-voce

*II **Phase** - GPA Calculation - To be done by St. Joseph's College (Autonomous)*

1. Consolidation of External and Internal for Theory Courses (Calculation of GPA)
2. Consolidation of External and Internal for Project Work (Calculation of GPA)
3. Consolidation of External and Internal for Practical Courses (Calculation of GPA)
4. Consolidation of External and Internal for Comprehensive Viva-voce (Calculation of GPA)

*III **Phase** - SGPA Calculation - To be done by St.Joseph's College (Autonomous) Irinjalakuda*

- Calculation of Semester Grade Point Average. This is the consolidated net result (Grade) in a particular Semester.

*IV **Phase** - CGPA Calculation - To be done by St.Joseph's College (Autonomous) Irinjalakuda*

- Calculation of Consolidated Grade Point Average. This is the consolidated net result (Grade) of a Programme.

Internal Evaluation / Continuous Assessment (CA)

Continuous Assessment will be based on a predetermined transparent system involving periodic two written tests, assignments, seminars and attendance in respect of theory courses and based on tests, lab skill and records/viva in respect of practical courses. The criteria and percentage of weightage assigned to various components for internal evaluation are as follows:

(a) Theory:			
Sl. No	**Component**	**Percentage**	**Weightage**
1	Examination /Test	40%	2
2	Seminars / Presentation	20%	1
3	Assignment	20%	1
4	Attendance	20%	1
(b) Practical:			
1	Lab Skill	40%	4
2	Records/viva	30%	3
3	Practical Test	30%	3

Attendance weightage 1 can be distributed as follows

Attendance	Internal weightage	Marks
Above 90%	1	5
85–89%	0.8	4
80–84%	0.6	3
76–79%	0.4	2
75%	0.2	1

Grades given for the internal evaluation are based on the grades A+, A, B, C, D & E with grade points 5, 4, 3, 2, 1 & 0 respectively. The overall grades will be as per the Ten Point scale. There shall be no separate minimum Grade Point for internal evaluation. To ensure transparency of the evaluation process, the internal assessment marks awarded to the students in each course in a semester will be published on the notice board before 5 days of commencement of external examination. There will not be any chance for improvement of internal marks. The course teacher will maintain the academic record of each student registered for the course.

***Examination /Test**:* For each course there shall be class test/s during a semester. Grades should be

displayed on the notice board. Valued answer scripts shall be made available to the students for perusal.

***Seminars / Presentation**:* Every student should deliver Seminar/Presentation as an internal built – in component of the curriculum transaction for every course and must be evaluated by the respective course teacher in terms of structure, content, presentation and interaction. The soft and hard copies of the seminar report are to be submitted to the course teacher.

Assignment: Each student will be required to do assignment/s as an internal built – in component of the curriculum transaction for each course. Assignments after valuation must be returned to the students. The teacher shall define the expected quality of the above in terms of structure, content, presentation etc. and inform the same to the students. Punctuality in submission is to be considered.

Lab Skill: Students in the science stream are required to combine their classroom methods with hands on practical sessions in the laboratories. The teacher shall assess the skills of the student and the quality of application of theoretical knowledge.

Records/viva: Records are submitted by science students for documenting the textual and classroom knowledge along with their practical lab skills. Neatness, accuracy and precision are also evaluated here. Viva voce is conducted to assess the grasp of knowledge gained by the student and to test their communication skills in the translation of the knowledge.

Practical Test: It is conducted for students in the science stream to assess their scientific temper and application of theoretical knowledge. The sense of precision and accuracy is also taken into account.

External / End Semester Evaluation (ESE)

The semester-end examinations in theory courses will be conducted by the Controller of Examination St. Joseph‟s College (Autonomous) Irinjalakuda with question papers set by external experts. The evaluation of the answer scripts will be done by examiners based on a well- defined scheme of valuation. The external evaluation will be done immediately after the internal valuation. The language of writing the examination should be English

Pattern of Questions For External/ESE:

Questions will be set to assess the knowledge acquired, standard, and application of knowledge, application of knowledge in new situations, critical evaluation of knowledge and the ability to synthesize knowledge. Due weightage will be given to each module based on content/teaching hours allotted to each

module. The question will be prepared in such a way that the answers can be awarded A+, A, B, C, D, E Grades. Different types of questions shall be given different weightages to quantify their range given in the following model:

Sl. No.	Type of Questions	Individual weightage	Total Weightage	Number of questions to be answered
1	Short Answer type questions	2	2 x 4 = 8	4 out of 7
2	Short essay/ problem solving type	3	3 x 4 = 12	4 out of 7
3	Long Essay type questions	5	5 x 2 = 10	2 out of 4
Total			**30**	**18**

End Semester Evaluation in Practical Courses will be conducted and evaluated by both Internal and External Examiners. (*Write about Duration and pattern of practical external examinations*)

Evaluation of project work / dissertation

There will be External and Internal evaluation with the same criteria for Project Work done and the grading system shall be followed. One component among the Project Work evaluation criteria will be Viva-voce (Project Work related) and the respective weightage will be 40%. Consolidated Grade for Project Work is calculated by combining both the External and Internal in the Ratio of 4:1 (80% & 20%). For a pass in Project Work, a student has to secure a minimum of P Grade in External and Internal examination combined. If the students could not secure minimum P Grade in the Project work, they will be treated as failed in that attempt and the students may be allowed to rework and resubmit the same in accordance with the University exam stipulations. There shall be no improvement chance for Project Work. The External and Internal evaluation of the Project Work shall be done based on the following criteria and weightages as detailed below:

Sl. No	Criteria	% of weightage	Weightage External	Weightage Internal
1	Relevance of the topic and Statement of problem	20%	8	2
2	Methodology & Analysis	20%	8	2
3	Quality of Report & Presentation	20%	8	2
4	Viva-Voce	(40%)	16	4
Total Weightage		**100%**	**40**	**10**

DIRECT GRADING SYSTEM

Direct Grading System based on a 10 – Point scale is used to evaluate the performance (External and Internal Examination of students). For all courses (Theory & Practical)/Semester/Overall Programme, Letter grades and **GPA/SGPA/CGPA** are given on the following way:

a) First Stage Evaluation for both Internal and External done by the Teachers concerned in the following Scale :

Grade	Grade Points
A+	**5**
A	**4**
B	**3**
C	**2**
D	**1**
E	**0**

b) The Grade Range for both Internal & External shall be :

Letter Grade	Grade Range	Range of Percentage (%)	Merit / Indicator
O	4.25 – 5.00	85.00 –100.00	Outstanding
A+	3.75 – 4.24	75.00 – 84.99	Excellent
A	3.25 – 3.74	65.00 – 74.99	Very Good
B+	2.75 – 3.24	55.00 – 64.99	Good
B	2.50 – 2.74	50.00 – 54.99	Above Average
C	2.25 – 2.49	45.00 – 49.99	Average
P	2.00 -2.24	40.00 – 44.99	Pass
F	< 2.00	Below 40	Fail
I	0	-	Incomplete
Ab	0	-	Absent

'B 'Grade lower limit is 50% and 'B+' Grade lower limit is 55%

No separate minimum is required for internal evaluation for a pass, but a minimum P Grade is required for a pass in the external evaluation. However, a minimum P grade is required for pass in a course. A student who fails to secure a minimum grade for a pass in a course will be permitted to write the examination along with the next batch.

Improvement of Course−The candidates who wish to improve the grade / grade point of the external examination of a course/s they have passed already can do the same by appearing in the external examination of the concerned semester along with the immediate junior batch.

Betterment Programme One time- A candidate will be permitted to improve the CGPA of the Programme within a continuous period of four semesters immediately following the completion of the

programme allowing only once for a particular semester. The CGPA for the betterment appearance will be computed based on the SGPA secured in the original or betterment appearance of each semester whichever is higher.

Semester Grade Point Average (SGPA) – Calculation

The SGPA is the ratio of sum of the product of the number of credits with the grade points scored by a student in all the courses taken by a student and the sum of the number of credits of all the courses taken by a student. After the successful completion of a semester, Semester Grade Point Average (SGPA) of a student in that semester is calculated using the formula given below.

Semester Grade Point Average - SGPA (S_j) = $\Sigma(C_i \times G_i)$ / Cr (SGPA= Total Credit Points awarded in a semester / Total credits of the semester)

Where „S_j„ is the j^{th} semester , „G_i „ is the grade point scored by the student in the i^{th} course 'c_i „ is the credit of the i^{th}course,'Cr ‟ is the total credits of the semester .

Cumulative Grade Point Average (CGPA) – Calculation

Cumulative Grade Point Average (CGPA) = $\Sigma(C_i \times S_i)$ / Cr(CGPA= Total Credit points awarded in all semesters/Total credits of the programme)

Where C_1 is the credit of the I^{st} semester S_1 is the SGPA of the I^{st} semester and Cr is the total number of credits in the programme. The CGPA is also calculated in the same manner taking into account all the courses undergone by a student over all the semesters of a programme. The SGPA and CGPA shall be rounded off to 2 decimal points. For the successful completion of a semester, a student should pass all courses and score a minimum SGPA of 2.0. However, the students are permitted to move to the next semester irrespective of their SGPA.

CONSOLIDATED SCHEME FOR I TO VI SEMESTERS

PROGRAMME STRUCTURE

STRUCTURE OF THE PROGRAMME

LEGEND	
Item	**Description**
C	Credits
E	External Component (%)
I	Internal Component (%)
L	Lecture Hours
P	Practical Hours
T	Total

No	Course Code	Course Name	C	Weightage			Hrs/Week		
				I	E	T	L	P	T
1.1	SJCSDC01	Principles of Data Science	4	1	4	5	4	0	4
1.2	SJCSDC02	Mathematics for Computing	4	1	4	5	3	2	5
1.3	SJCSDC03	Data Mining	4	1	4	5	4	0	4
1.4	SJCSDC04	Python Programming	4	1	4	5	2	2	4
1.5	SJCSDC05	Design and Analysis of Algorithms and Object	4	1	4	5	4	0	4
1.6	SJCSDL01	Python and C++ practical	2	1	4	5	0	4	4
1.7	SJCSDA01	Introduction to Research (Ability Enhancement Audit Course)	4	5	0	5	0	0	0
		Total Credits(Excluding Audit Course):22					**17**	**8**	**25**

SEMESTER I

SEMESTER II

No	Course Code	Course Name	C	Weightage			Hrs/Week		
				I	E	T	L	P	T
2.1	SJCSDC06	Advanced Database Management	4	1	4	5	4	0	4
2.2	SJCSDC07	Operations Research	4	1	4	5	3	2	5
2.3	SJCSDC08	R Programming for Data Analytics	4	1	4	5	4	0	4
2.4	SJCSDC9	Data Visualization	4	1	4	5	2	2	4
2.5	SJCSDC10	Linear Algebra for Machine Learning	4	1	4	5	4	0	4
2.6	SJCSDL02	R programming and ADBMS practical	2	1	4	5	0	4	4
2.7	SJCSDA02	Term Paper (Professional Competency Audit Course)	4	5	0	5	0	0	0
Total Credits(Excluding Audit Course):22							**17**	**8**	**25**

		SEMESTER III							
No	Course Code	Course Name	C	Weightage			Hrs/Week		
				I	E	T	L	P	T
3.1	SJCSDC11	Cloud Computing	4	1	4	5	3	1	4
3.2	SJCSDC12	Map Reduce Programming	4	1	4	5	2	3	5
3.3	SJCSDC13	Machine Learning	4	1	4	5	4	0	4
3.4	SJCSDE01	Elective I	4	1	4	5	4	0	4
3.5	SJCSDE02	Elective 2	4	1	4	5	4	0	4
3.6	SJCSDL03	Map Reducing and Machine learning practical	2	1	4	5	0	4	4
		Total Credits(Excluding Audit Course):22					**17**	**8**	**25**

SEMESTER IV

No	Course Code	Course Name	C	Weightage			Hrs/Week		
				I	E	T	L	P	T
4.1	SJCSDE03	Elective 3	3	1	4	5	5	0	5
4.2	SJCSDE04	Elective 4	3	1	4	5	5	0	5
4.3	SJCSDP01	Project Requirements Analysis & Design							
		Related Discussion					3	1	4
		Project Coding, Testing & Implementation	8	1	4	5	2	2	4
		Related Discussion							
		Project Evaluation & Assessment					2	0	2
		Project Lab Work/Internship					0	5	5
		Total Credits(Excluding Audit Course):14					**17**	**8**	**25**

List of Elective Courses

Course Code	Course Name
SJCSDE01	Text Analytics
SJCSDE02	Sentiment Analysis
SJCSDE03	Online Course I
SJCSDE04	Online Course II

SEMESTER I

Course Code: SJCSDC01

Name of the Course: Principles of Data Science

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)
CO1	Understand Data sources, generations, data formats, DataEvolution, Data from various domains	PSO2	R, U	F	10
CO2	Understand Big Data Characteristics What, Why, When,Limitation of traditional approaches and models. Map BigVs to Data Domains	PSO2, PSO3	A	C	10
CO3	Understand Big Data Processing platform, frameworks,Hadoop, Spark , storage models – Hbase	PSO1	Z	P	10
CO4	Programming Model of Big Data MapReduce, WhyMapReduce, Limitations of Traditional Models	PSO1	A	P	10
CO5	Analyze various domains of Big Data Characteristics,Platform, Programming Model	PSO4	Z	P	10
CO6	Design Big Data framework ecosystem, and data processingframework of multidisciplinary domains	PSO3	Z	C	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : PRINCIPLES OF DATA SCIENCE

Course Code :SJCSDC01

Course Objectives

To impart knowledge to make the students

1. To understand Data source evolution, data Characteristics and data processing models.
2. To understand and apply data process architecture HADOOP, SPARK and write MapReduce programs.
3. To analyze Big Data use cases for specific domain and applications.

UNIT I

Data Evolution: Data Development Time Line – ICT Advancement-a Perspective – Data Growth-a Perspective– IT Components-Business Process – Landscape-Data to Data Science – Understanding data: Introduction –Type of Data: Numeric – Categorical – Graphical – High Dimensional Data — Data Classification – Hot Data –Cold Data – Warm Data – Thick Data – Thin Data - Classification of digital Data: Structured, Semi-Structuredand Un-Structured. Sources of Data: Time Series – Transactional Data – Biological Data – Spatial Data – Social Network Data – Data Evolution – Data Sources

UNIT II

Data Science: Data Science-A Discipline – Data Science vs Statistics – Mathematics - Programming Language -Database, - Machine Learning. Data Analytics Relation: Data Science, Analytics, Big Data Analytics. DataScience Components: Data Engineering, Data Analytics-Methods and Algorithm, Data Visualization Big Data:Introduction To Big Data: - Evolution What is Big Data – Sources of Big Data. Characteristics of Big Data 6Vs– Big data-Challenges of Conventional Systems- -– Data Processing Models – Limitation of Conventional Data Processing Approaches – Big Data Myths - Data Discovery-Traditional Approach, Big Data Technology: Big Data Exploration - Data Augmentation – Operational Analysis – 360 View of Customers – Security and Intelligence

UNIT III

Hadoop: Basic Concepts-An Overview of Hadoop-The HadoopDistributed File System-Anatomy of a Hadoop Cluster-Hadoop Ecosystem Components. Replica-Hadoop Processes-Name node-Secondary name node-Job tracker-Task tracker-Data node – Hadoop YARN – Hadoop Limitation - SPARK – in Architecture – SPARK Advantages - HBASE: HBase Architecture-HBase API-Managing large data sets with HBase

UNIT IV

Map Reduce: Developing Map Reduce Application - Phases in Map Reduce Framework - Map Reduce Input and Output Formats - Advanced Concepts - Sample Applications – Combiner – Joining datasets in Mapreduce jobs – Map - side join – Reduce - Side join - Map reduce – customization - **Map Reduce Program:** Introduction to Writing a MapReduce Program - The MapReduce Flow - Examining a Sample MapReduce Program- Basic MapReduce API Concepts - The Driver Code - The Mapper - The Reducer - Example

UNIT V

Big Data Usecases –Big Data Technology Potentials – Limitations of Big Data and Challenges- Big Data Roles Data Scientist , Data Architect, Data Analyst – Skills – Case Study : Big Data – Customer Insights – Behavioral Analysis – Big Data Applications - Marketing – Retails – Insurance – Risk and Security – Health care

REFERENCE

1. V. Bhuvaneswari, T. Devi, "Big Data Analytics: A Practitioner"s Approach" 2016.
2. Han Hu, Yonggang Wen, Tat-Seng, Chua, XuelongLi,"Toward Scalable Systems for Big
3. SeemaAcharya, SubhashniChellappan, "Big Data Analytics", Wiley, 2015.
4. Han Hu, Yonggang Wen, Tat-Seng, Chua, XuelongLi,"Toward Scalable Systems for Big Data Analytics: A Technology Tutorial", IEEE, 2014

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS –
SEMESTER 1
MODEL QUESTION PAPER
SJCSDC01 PRINCIPLES OF DATA SCIENCE

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightage.*

1. Explain Data Evolution and Data Growth
2. Explain Type of Data
3. Describe Data Classification
4. Influence of Data Science Components in analysis
5. Characteristics of Big Data
6. Explain anyone application of Big Data PAR
7. Explain the need of Data Analytics with a simple scenario

Part B

Answer **any four** *questions.*

*Each question carries **3** weightages.*

8. Explain 360 view of customers on the perspective of big data.
9. Illustrate Hadoop ecosystem components.
10. Explain SPARK architecture and its advantages.
11. Brief about phases in Map Reduce framework.
12. What is the basic API concept of Map Reduce?
13. Explain the different big data role.
14. Explain the limitations of big data.

Part C

Answer **any two** *questions.*

*Each question carries **5** weightages.*

15. Explain the different sources of data with example.
16. Explain the Hadoop architecture and its process.
17. Brief the influencing of big data in Risk and security domain.
18. Explain the progress of big data technology in real time applications.

Course Code: SJCSDC02

Name of the Course: Mathematics of Computing

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**
CO1	Understand the principles of probability, frequency distribution measures.	PSO1	U	C	10

CO2	Understand the correlation and regression, hypothesis test, sampling techniques for specific applications.	PSO1	A	C	10
CO3	Apply probabilistic models and distribution models	PSO5	A	P	10
CO4	Apply hypothesis testing and regression models for specific domain	PSO5	Z	P	10
CO5	Illustrate statical model and Infer	PSO1	E	F	10
CO6	Design statistical models for specific domains.	PSO5	C	F	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : MATHEMATICS OF COMPUTING

Course Code :SJCSDC02

Course Objectives

To impart knowledge to make the students

1. To understand the Probability Theory
2. To understand regression and theoretical distributions.

UNIT I

Measures of Central Tendency: Arithmetic Mean, Geometric Mean - Harmonic Mean-Median, Mode - Dispersion: Overview - Mean Deviation - Standard Deviation - Combined Standard Deviation.

UNIT II

Basic probability theory - distributions and their properties - Frequency Distribution - Continuous or Grouped Frequency Distribution - Magnitude of Class intervals - Cumulative Frequency Distribution - Two Way Frequency Distribution

UNIT III

Regression: Overview - Simple and multiple regression analysis - Regression, Graphical Method - Algebraic Method - Regression Line - Regression Equation, - hypothesis testing - Hypothesis - Standard Error - Test of Significance for Attributes - Test of Significance for Large Samples - Test of Significance for Small Samples - Chi Square Test - sampling - estimation theory - least square methods - SVD - transformations

UNIT IV

Theoretical Distribution: Binominal Distribution - Obtaining Coefficient - Poison Distribution - Normal Distribution - Poisson - Cumulative Poisson Process and its generalization - applications in different business domain

UNIT V

TIME SERIES:- Introduction-Components of time series- time series models(MA,AR,ARMA,ARIMA,SARIMA,LSTM)- Univariate and multivariate time series models -Time series data –Autocorrelation-Time series applications, Stochastic models compression techniques - Markov Models - Markov decision process - application in sequential decision making - Monte Carlo Simulations

REFERENCES

1. R.S.N. Pillai, Bagavathi, “Statistics Theory and Practice, S.Chand& Company, 2013
2. Douglas C. Montgomery, George C. Runger., “Applied Statistics for Engineers”, John Wiley &Sons.Inc, 2003

Model Question Paper

ST. JOSEPH’S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS –
SEMESTER 1
MODEL QUESTION PAPER
SJCSDC02 MATHEMATICS FOR COMPUTING

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightage*

1. In a moderately asymmetrical distribution Mean is 24.6 and Median 25.1. Find the value of mode.
2. Calculate GM and HM for the following data

Value	0-10	10-20	20-30	30-40	40-50
Frequency	8	12	20	6	4

3. (a)Find the arithmetic mean of the following frequency distribution:

x	1	2	3	4	5	6	7
f	5	9	12	17	14	10	6

(b)Calculate the arithmetic mean of the marks from the following table:

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No.of.Students	12	18	27	20	17	6

4. In a random arrangement of the letters of the word 'COMMERCE', find the probability that all the vowels together?
5. Four cards are drawn at random from a pack of 52 cards. Find the probability that
   (i)They are a king, a queen, a jack and ace
   (ii)Two are kings and two are queens
6. Explain the commonly used Measures of dispersion.
7. Explain ARMA.

Part B

Answer **any four** *questions.*

*Each question carries **3** weightages.*

8. The king, queen and jack of clubs are removed from a deck of 52 playing cards and then shuffled. A card is drawn from the remaining cards. Find the probability of getting:

   (i) a heart

   (ii) a queen

   (iii) a club

   (iv) '9' of red colour

9. The amount that the children have spent for purchasing some eatables in one day trip of a school are :5, 10, 15, 20, 25, 30, 35, 40. Using step deviation method, find the standard deviation of the amount they have spent.

10. The mean number of close friends for the population of people living in the U.S. is 5.7. An investigator predicts that the mean number of close friends for introverts will be significantly less than the mean of the population. The mean number of close friends for a sample of 26 introverts

is 6.5. The standard deviation of scores in this sample is 1.3. Do these data support the investigator's prediction? Use an alpha level of .05.

11. A genetics engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio 4 stripes only: 3 spots only: 9 both stripes and spots. When the cross was performed and she counted the individuals she found 50 with stripes only, 41 with spots only and 85 with both. According to the Chi-square test, did she get the predicted outcome?

12. Explain Markov Decision Process.

13. Derive Mean, variance and mgf of binomial distribution.

14. X is a normally normally distributed variable with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find
    a) $P(x < 40)$
    b) $P(x > 21)$
    c) $P(30 < x < 35)$

Part C

Answer **any two** *questions.*

Each question carries ***5*** *weightages.*

15. Explain Measures of Central Tendency with Merits, Demerits and Example.
16. Explain the CLT.
17. Explain ARIMA and AROMA.
18. Explain the application of markov decision process.

Course Code: SJCSDC03

Name of the Course: Data Mining

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**
CO1	Understand data mining tools and techniques and big data for various domains.	PSO1	R,U	C	10

CO2	Apply various data mining, text mining, web mining algorithms for real time applications.	PSO5	A	P	10
CO3	Analyze unsupervised and supervised algorithms for real world applications.	PSO5	Z	P	10
CO4	Illustrate the mining techniques like association, classifications and clustering on datasets	PSO1	E	P	10
CO5	Apply R programming packages for mining data.	PSO1	A	P	10
CO6	Compare various approaches of data mining algorithms.	PSO3	E	P	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : DATA MINING

Course Code :SJCSDC03

Course Objectives

To impart knowledge to make the students

1. To understand the concepts of Data Warehouse architecture and apply for various domains.

2. Tounderstand DataMining techniques Cluster, Classification and Association Rule Mining.

3. To understand the concepts of Web mining, Text mining and Spatial mining.

UNIT I

Data warehousing: Introduction - Definition - Multidimensional data model - OLAP operations - Warehouse schema - Data warehousing architecture - Warehouse Schema - Warehouse server - Meta data - OLAP Engine - Data warehouse backend process - Data Warehouse Technology - Warehousing

Software - Cloud data warehousing - Other features. Data Warehousing Case Study: Government, Tourism and Industry

UNIT II

Data mining: Introduction – Data as a Subject - Definitions- KDD vs. Data mining- DM techniques-CurrentTrends in Data Mining. Association Rules: Concepts- Methods to discover Association rules- A priori algorithm– Partition algorithm- Pioneer search algorithm –Dynamic Item set Counting algorithm- FP-tree growth algorithm-Incremental algorithm-Border algorithm-Generalized association rule. Analysis of association rule using orange.

UNIT III

Clustering techniques: Data Attribute Types – Data Similarity and Dissimilarity - Clustering paradigms– Partition algorithm-K- Medeoid algorithms – CLARA- CLARANS –Hierarchical DBSCAN-BIRCH- CURE-Categorical clustering algorithms-STIRR-ROCK-CACTUS-Other techniques: Implementation of Clustering techniques using orange tool.

UNIT IV

Classification Technique: Introduction – Decision Trees: Tree Construction Principle – Attributte Selection measure – Tree Pruning - Decision Tree construction Algorithm – CART – ID3 - Rainforest - CLOUDS - BOAT, Pruning Technique – Model Evaluation –Cross Validation – Bootstrap – Holdout – Classifier Performance- Boosting – AdaBoost - Bagging

UNIT V

Web mining: Basic concepts – Web content mining – Web structure mining – Web usage mining – text mining– Text Preprocessing - Text clustering – Spatial mining – Spatial mining tasks – Spatial clustering – Spatial trends – Case Studies : Big Data , IoT..

REFERENCE

1. Jiawei Han, MichelineKamber, “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, 2012
2. Arun K Pujari, “Data Mining Techniques”, Universities Press. 2012

Model Question Paper

ST. JOSEPH’S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS –
SEMESTER 1
MODEL QUESTION PAPER
SJCSDC03 DATA MINING

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightages*

1. Explain Data Warehouse Architecture with diagram

2. Explain operations of OLAP

3. Describe about the common sectors where Data Warehouse is used (at least 4) 4. What are the steps for KDD process?

5. Data Mining techniques and their descriptions

6. Use cases for association rules (at least 4)

7. What do you understand by web content mining?

Part B

Answer **any four** *questions.*

*Each question carries **3** weightages.*

8. Explain cloud data warehousing on the basis of case study

9. Describe data warehouse technology and warehousing software.

10. Differentiate between KDD Vs Data mining.

11. Explain the current trend in data mining.

12. Explain about Hierarchical DBSCAN, BIRCH, CURE.

13. Explain Decision tree with an example.

14. What is web structure mining?

Part C

Answer **any two** *questions.*

*Each question carries **5** weightages*

15. Explain pioneer search algorithm.

16. Explain the influence of implementation of clustering techniques
using orange tool.

17. Briefly explain boosting, adaboost and bagging

18. Compare web content mining, text mining and spatial mining.

Course Code: SJCSDC04

Name of the Course: Python Programming

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)	Lab(Hrs)
CO1	Understand the object-oriented concepts and control structures.	PSO1	U	F	6	4
CO2	Understand the basic programming structure-list, dictionary, tuple.	PSO1	A	P	6	4
CO3	Apply OOPs concept for designing software applications.	PSO5	A	P	6	4
CO4	Understand the visualization methods, packages, statistical packages and other packages for building data models	PSO4	E	C	6	4
CO5	Design and analyze dataset applying statistical models, visualization and models using various tools.	PSO5	Z	P	6	4
CO6	Design data analytic model using the packages in python and provide inference for multidisciplinary domains.	PSO4	C,E	P	6	4

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : PYTHON PROGRAMMING

Course Code :SJCSDC04

Course Objectives

To impart knowledge to make the students

1. To understand the basics of Python Data structures and Programming constructs.
2. To understand and Apply Python Libraries for Data Science and Machine Learning

UNIT I

INTRODUTION

Introduction to Python: Python Introduction, History of Python, Python features , Python interpreter, Overview of programming in Python, Basic data types Python built in types, Arithmetic in Python, Program input and Program output, Variables and assignment. Global and local variables.Modules :Importing module, Math module Random module, Packages, Composition. Exception Handling .

UNIT II

ADVANCED DATA TYPES

Python Strings and string manipulation [Assigning values in strings, String manipulations, String special operators, String formatting operators, Triple Quotes, Raw String, Unicode String, Build-in-String methods], Python List : Introduction, Accessing values in list, List manipulations, List Operations, Indexing, slicing & matrices. Python Dictionary - Introduction, Accessing values, Properties, Functions in Dictionary. Python Tuples : Introduction, Operation, Accessing , Function and methods in tuples andData Type Conversion.

UNIT III

CONTROL STRUCTURES

Python - Basic Operators: Arithmetic Operators ,Comparison Operators, Logical (or Relational) Operators,Assignment Operators, Conditional (or ternary) Operators Conditional Statement : Branching (if, else-if,nested),Looping : while statement, for statements, Control Statements: break, continue and pass Statements.Functions : Defining a function , Calling a function ,Types of functions , Function Arguments Anonymousfunctions , Regular expressions : Match function,Search function ,Modifiers. OOPs concept.

UNIT IV

PYTHON LIBRARIES FOR DATA SCIENCE

NumPy [Arrays and matrices]: N-dimensional data structure, Creating array, Indexing array, Reshaping,Vectorized operations, Pandas [Data Manipulation]: Create Data Frame, Combining Data Frames, Summarizing, Columns selection, Rows selection (basic) , Rows selection (filtering) , Sorting, Descript ive statistics, Rename values, Dealing with outliers

UNIT V

STATISTICS AND MACHINE LEARNING IN PYTHON

SciPy Introduction, Basic functions, Special functions(scipy.special), Integration(scipy.integrate), Optimization (scipy.optimize), Visualization libraries : matplotlib, Seabor. Univariate Statistics, Dimension Reduction And Feature Extraction, Classification - Case Studies

REFERENCES

1. Core Python Programming by Wesley J. Chun, 2nd Edition ,Pearson Education
2. An Introduction to Python by Guido Van Russom, Fred L.Drake, Network Theory
3. Limited.
4. Beginning Python: From Novice To Professional By Magnus Lie Hetland, Second Edition.
5. Programming in Python 3 by Mark Summerfield, Pearson Education
6. Python for Probability, Statistics, and Machine Learning, by Unpingco J.
7. Statistics and Machine Learning in Python, by EdouardDuchesnay, Tommy Löfsted

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS –
SEMESTER 1
MODEL QUESTION PAPER
SJCSDC04 PYTHON PROGRAMMING

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

Each question carries ***2*** *weightages*

1. Briefly explain the features in python.(minimum 6)
2. Explain the Standard Data Types in python.
3. Explain Global and Local variables with example.
4. What are packages?
5. Explain NumPy, Pandas and OpenCV packages in python.
6. Define String. Explain how to access values in a string, update string and string concatenation (with examples).
7. Mention and explain List Methods.

Part B

Answer **any four** *questions.*

Each question carries ***3*** *weightages*

8. Illustrate the different types of control flow statements available in Python.
9. Explain mathematical functions, date time functions.
10. Differentiate Simple if statement, if-elif-else statement with an example.
11. Write Python Program to reverse a number
12. Explain the concept of scope and lifetime of variables in Python programming.
13. Differentiate identifiers and keywords.
14. Explain built in functions.

Part C

Answer **any two** *questions.*

Each question carries ***5*** *weightages*

15. Explain while loop and for loop with suitable example.
16. Explain Strings and string operations.
17. Explain Dictionary and Tuples. Explain the below with an example:

a. Declaring a dictionary (Generalized form)
b. Accessing values in dictionary and Tuples
c. Built-in Functions in Dictionary and Tuples.
d. Negative Indexing in tuples.
e. Slicing in list.
f. String Special Operators.

18. Explain pandas and data manipulation included in it.

Course Code: SJCSDC05

Name of the Course: Design and Analysis of Algorithms and Object

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**	**Lab(Hrs)**
CO1	Identify classes and objects from the given problem description and create classes and objects using C++	PSO1	U	C	5	5
CO2	code reusability and extensibility by means of Inheritance and Polymorphism	PSO5	A	P	5	5
CO3	Differentiate among various algorithmic approaches	PSO5	A	P	5	5

CO4	Design algorithms for problem solving by using the suitable algorithmic technique	PSO5	A	P	5	5
CO5	Analyze a given algorithm for its efficiency based on time and space it occupies	PSO5	A	P	5	5
CO6	Apply optimization techniques for improving the performance of algorithms.	PSO3	Z	P	5	5

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : Design and Analysis of Algorithms and Object

Course Code :SJCSDC05

Course Objectives

To impart knowledge to make the students

1. To understand the object oriented concepts: Class, Inheritance and Polymorphism.

2.To understand and analysis concepts of Algorithmic analysis and algorithm approaches.

UNT I

Object oriented language fundamentals – programming basics – Conditional statements – Structures – Functions - Objects and Classes – Constructors – Overloading.

UNIT II

Inheritance – Hierarchy - Derived class – Access specification - Polymorphism – virtual functions – virtual class – Files - Exception Handling.

UNIT III

Introduction to algorithms, Analyzing algorithms. Divide and Conquer: General Method, Binary Search, Merge sort, Quick sort.

UNIT IV

Greedy Method: Knapsack problem, Job sequencing with deadlines, Minimum spanning trees, Single source shortest paths. Dynamic Programming: Multistage graphs, All pairs shortest paths, Travelling salesperson problem.

UNIT V

Back Tracking: 8-queens problem, Sum of subsets, Graph coloring, Hamiltonian cycles. Branch and Bound:General method, Travelling salesperson problem.

REFERENCES

1. BjarneStroustrup, "The C++ Programming Language", Addison Wesley, 4th Edition, 2013.

2. Ellis Horowitz, SartajSahni, SanguthevarRajasekaran, "Fundamentals of Computer Algorithms", Galgotia Publications, 2011.

3. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, "Introduction to Algorithms", Massachusetts Institute of Technology, MIT Press, III Edition, 2009.

4. M.A.Weiss, "Data Structures and Algorithm Analysis in C++", Pearson Education Asia,2013.

5. www.spoken-tutorial.org

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS –
SEMESTER 1
MODEL QUESTION PAPER
SJCSDC05 OBJECT ORIENTED PROGRAMMING CONCEPTS

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

Each question carries ***2*** *weightages*

1. Explain the important features of Object Oriented Programming.
2. Explain the types of constructors in C++.
3. Explain the 6 conditional statements in C++.
4. Explain structure and give the general syntax.

a. How to declare structure variables
b. How to initialize structure members

5.Difference between Call by Value and Call by Reference with examples.
6. Define Inheritance and write the syntax. Mention and explain the types of Inheritance.
7. What is function overloading?

Part B

Answer **any four** *questions.*

Each question carries ***3*** *weightages*

8. Write an algorithm for merge sort.
9. Explain binary search with suitable example.
10. Explain minimum spanning trees steps with example.
11. What is an algorithm? How to analyze algorithm.
12. Explain Hamiltonian cycles.
13. Explain polymorphism concept with example.
14. What is file? Explain exception handling.

Part C

Answer **any two** *questions.*

Each question carries ***5*** *weightages*

15. Explain the two types of polymorphism.
16. Define Virtual Function and Pure virtual function. Give an example for virtual function.
17. Explain travelling salesman problem with respect to branch and bound.
18. Explain knapsack problem with an example.

Course Code: SJCSDL01

Name of the Course: Lab1: Python and C++

	Course Outcome	POs / PSOs	CL	KC	Lab(Hrs)
CO1	Implement Classes and Objects, Constructors and Destructors with array of Objects, Passing and returning parameters as objects by reference, String manipulation functions, different types of inheritances like Multiple, Multilevel and Hybrid.	PSO1	A	P	5
CO2	Demonstrate Function Overloading, overload different operators – incr and decr operators with post and pre forms, use of Virtual Functions	PSO1	A	P	5
CO3	.Read, write, and execute simple Python programs, write simple Python programs for solving problems	PSO5	A	P	5
CO4	Decompose a Python program into functions, lists etc.	PSO4	A	P	5
CO5	Read and write data from/to files in Python Programs	PSO5	A	P	5
CO6	Underline the use of package	PSO4	A	P	5

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Part A - Python Programming Practical

- Advanced Data Types- List,Tuple,Dictionary,String

- Control Structures- Looping statements, Conditional statements
- Functions
- OOPS Concept- Classes and Objects, Inheritance
- Data Analytics using Basic Python
- Data Analytics using NUMPY
- Data Analytics using PANDAS
- Data Analytics using MATPLOTLIB

Part B – C++ Practical

C++_ Programs

- Student details using CLASS and OBJECT
- Product of two numbers using Single Inheritance
- Electricity Bill calculation using Multilevel Inheritance
- Virtual Function to display the numbers
- Function overloading to calculate the area of shapes
- Add complex numbers using Operator overloading
- Bank details using Multiple Inheritance

Quantitative Aptitude

- Encoding and Decoding
- Blood Relation
- Logical Venn Diagram
- Mathematical operations
- Arithmetic Reasoning
- Truth of the statements
- Missing figures
- Detecting Incorrect order
- Detecting the wrong figures

SEMESTER II

Course Code: SJCSDC06

Name of the Course: Advanced Database Management

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**	**Lab(Hrs)**
CO1	Explain the structure and model of the relational database system	PSO1	U	C	10	5
CO2	Design multiple tables and using group queries	PSO5	A	P	8	7
CO3	Design a database based on a data model normalization to a specified level	PSO5	Z	P	8	7
CO4	Mongo DB and Operators	PSO5	A	P	8	7

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : ADVANCED DATABASE MANAGEMENT

Course Code :SJCSDC06

Course Objectives

To impart knowledge to make the students

1) To understand the concepts of DBMS, Data Model and Normal forms.
2) To understand the concepts of concurrency control and Recovery.
3) To understand basics of SQL and NoSQL databases.
4) To understand and apply MongoDB (NoSQL) for Data Analysis using CURD and User Management.

UNIT I

Introduction - Database concepts, Basic components of DBMS, sources of data - data models – hierarchical – network – XML and Stores - Relational Database Design: Anomalies ina Database–Functional Dependency – Lossless Join and Dependency – Preserving Decomposition – Third Normal Form– BoyceCodd Normal Form – Multivalued Dependency – Fourth Normal Form – Join Dependency – Project Join Normal Form –Domain Key Normal Form - SQL: Data Definition – Data Manipulation – Integrity Constraints–Views–PL/SQL.

UNIT II

Indexing and Hashing – Query Processing – Transaction Processing – Concurrency Control and Recovery - Advanced Database Concepts and Emerging Applications: Distributed Databases – Object Oriented Databases - Object Relational Databases- Data mining and Data Warehousing – Big Data - Big Databases- SQL–NoSQL Tradeoffs–CAP Theorem–Eventual Consistency - NoSQL–database types – Document Oriented – Columnar – Graph – KeyValue Pair - NoSQL database, design for performance / quality parameters, documents and information retrieval

UNIT III

MongoDB- Introduction - MongoDb – Need – MongoDBVs RDBMS – MongoDB- Driver Installation – Configuration – Import and Export – MongoDB Server Configuration - Data Extraction Fundamentals - Intro to Tabular Formats - Parsing CSV -Parsing XLS with XLRD-Parsing XML - Intro to JSON - Getting Data into MongoDB - MongoDB- CURD – Database Creation – Update – Read – Delete

UNIT IV

Using mongoimport -Operators like $gt, $lt, $exists, $regex -Querying Arrays and using $in and $all Operators -Changing entries: $update, $set, $unset - Data Analysis - Field Queries -Projection Queries-

Limiting – Sorting- Aggregation - Examples of Aggregation Framework -The Aggregation Pipeline - Aggregation Operators: $match, $project, $unwind, $group

UNIT V

User Management – MongoDb Data Replication in Servers – Data Sharding – MongoDB Indexes – Create – Find – Drop – Backup – MongoDB – Relationships – Analyzing Queries – MongoDBObjectid - Advanced MongoDB:MapReduce – MongoDB - Text Processing - Regular Expression – Case Studies – Text processing of large datasets, Map Reduce using MongoDB

REFERENCES

1. Abraham Silberchatz, Henry K.Forth, Sudharshan, "Database system Concepts" – (6th edition), McGraw Hill, 2010.
2. Elisa Bertino, "Object Oriented Databases", Addison Wesley. 1993. 3. RamezElmasri,
3. ShamkantB.Navathe, " Fundamentals of Database Systems ", 3rd Edition, Addison Wesley-2000. 4. Malay
4. k. Pakhira, "Database Management System", Phi Learning Pvt. Ltd., 2012
5. www.spoken-tutorial.org
6. MongoDB: The Definitive Guide, 2nd Edition , Powerful and Scalable Data Storage,

   By Kristina Chodorow, Publisher: O'Reilly Media

7. MongoDB Basics - EelDavid Hows,Peter Membrey,coPlugge, Publisher Apress - Ebook(free) https://it-ebooks.info/book/4527/

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS

SEMESTER II MODEL QUESTION PAPER

SJCSDC06 ADVANCE DATABASE MANAGEMENT

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

Each question carries ***2*** *weightage.*

1. What is meant by data abstraction?
2. Explain derived attribute with a example.
3. List the DDL commands in SQL.
4. What are the different types of anomalies that may occur while designing a database?
5. Explain the various keys in DBMS.
6. **What is Data Redundancy?**
7. **What are the main tasks performed by DBA?**

Part B

Answer **any four** *questions.*

Each question carries ***3*** *weightages.*

8. Explain Big Data and Big Databases.
9. Explain Integrity constraints.
10. Explain normalization. Mention the normal forms.
11. List DML commands in SQL
12. Write the advantages of Mongo DB over DBMS
13. Explain the aggregation Operators.
14. Explain the basic components of DBMS.

Part C

Answer **any two** *questions.*

Each question carries ***5*** *weightages*

15. Explain the NoSQL database types.
16. Explain the applications of DBMS in real life
17. Brief about advanced database concepts and emerging applications.
18. Explain a case study on Map Reduce using MongoDB with program.

Course Code: SJCSDC07

Name of the Course: Operation Research

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**
CO1	Identify the Assignment problem and to optimize in engineering fields	PSO1	U	C	10
CO2	Solve linear programming techniques to optimization problems	PSO5	A	P	10

	arising in all Computer fields				
CO3	Solve Integer linear programming techniques to optimization problems arising in all Computer fields	PSO5	A	P	10
CO4	Use Dynamic programming approach to real time problems	PSO3	A	P	10
CO5	Compare different Transportation algorithms	PSO3	U	C	10
CO6	Write a case study using any Operations Research methods to get optimal solution for an organization	PSO5	Z	P	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : OPERATION RESEARCH

Course Code :SJCSDC07

Course Objectives

To impart knowledge to make the students

1. To understand linear programming methods
2. To understand Dynamic programming approach

UNIT I

Introduction to Operations Research: Basics definition - scope – objectives - phases - models - limitations of Operations Research - Linear Programming Problem - Formulation of LPP - Graphical solution of LPP - Simplex Method - Artificial variables - big-M method - two-phase method -

degeneracy - nbound solutions - Introduction to optimization - gradient descent method - convex optimization.

UNIT II

BIG-M method – Two – Phase method – Special cases in the Simplex method – Transportation and Assignment Problems – Revised Simplex Method – Duality in Linear Programming Problems – Dual Simplex method – Bounded variable technique – Integer programming: Knapsack Problem – Cutting plane algorithm – Branch and bound programming – Mixed integer Programming – travelling salesperson problem.

UNIT III

Dynamic programming: Dynamic programming. Characteristics of dynamic programming – Dynamic programming approach for Priority Management employment smoothening – capital budgeting – Stage Coach/Shortest Path – cargo loading and Reliability problems.

UNIT IV

Assignment problem: Formulation – Hungarian method for optimal solution - Solving unbalanced problem – Traveling salesman problem and assignment problem.

UNIT V

Transportation problem and their applications: Transportation Problem – Formulation, solution – unbalanced Transportation problem – Finding basic feasible solutions – Northwest corner rule – least cost method and Vogel‟s approximation method. Optimality test: the stepping stone method and MODI method.

REFERENCE

1. J K Sharma, “Operations Research Theory &Applications , 3e”,Macmillan IndiaLtd,2007
2. P. K. Gupta and D. S. Hira, “Operations Research”, S. Chand & co., 2007.
3. J K Sharma., “Operations Research, Problems and Solutions, 3e”, Macmillan India Ltd.
4. N.V.S. Raju, “Operations Research”, HI-TECH, 2002

Model Question Paper

ST. JOSEPH’S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS –
SEMESTER 1II

MODEL QUESTION PAPER
SJCSDC07 OPERATION RESEARCH

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightage.*

1. Explain the slack and surplus variable with example
2. Solve the following LPP graphically

   Minimise $Z = 5x_1 + 8x_2$

   Subject to $6x_1 + 2x_2 \geq 12$

   $2x_1 + 2x_2 \geq 8$

   $4x_1 + 12x_2 \geq 24$

   $x_1, x_2 \geq 0$
3. Explain integer programming knapsack problem
4. Explain Hungarian method
5. What is unbalanced transportation problem?
6. Explain Big M method
7. What are the applications and limitations of Operations research?

Part B

Answer **any four** *questions.*

*Each question carries **3** weightages.*

8. Explain NWCM and find the initial basic feasible solution of the given transportation problem using NWCM

	D_1	D_2	D_3	D_4	Supply
O_1	1	2	1	4	30
O_2	3	3	2	1	50
O_3	4	2	5	9	20
Demand	20	40	30	10	

9. Compare Big M and Two phase methods
10. A company produces 3 products A, B and C from three raw materials P, Q and R. One unit of product A requires 4 units of P and 6 units of Q. One unit of product B requires 4 units Q and10 units of R and one unit of Product C requires 6 units of P and 4 units of Q and 8 units of R. The company has 16 units of material P,20 units of Q and 30 units of R. Profit per unit of Products A, B and C are $Rs.6$,$Rs.10$ and $Rs.8$ respectively.

    Formulate the problem mathematically to maximize profit

11. Explain canonical and standard form of LPP using example
12. Solve the following LPP graphically

    Minimise $Z = 5x_1 + 8x_2$

    Subject to $6x_1 + 2x_2 \geq 12$

    $2x_1 + 2x_2 \geq 8$

    $4x_1 + 12x_2 \geq 24$

    $x_1, x_2 \geq 0$

13. Solve the transportation problem by Least Cost Method

	D_1	D_2	D_3	D_4	Supply
O_1	2	3	11	7	6
O_2	1	0	6	1	1
O_3	5	8	15	9	10
Demand	7	5	3	2	17

14. Explain cutting plane algorithm

Part C

Answer **any two** *questions.*

Each question carries ***5*** *weightages*

15. What is assignment problem and solve the following assignment problem

Computer operators	Task			
	A	B	C	D
P	17	10	6	8
Q	18	8	14	9
R	14	12	12	10
S	15	9	7	11

16. Solve the linear programming problem using Two phase method

Max $Z = 3x_1 - x_2$

Subject to $2x_1 + x_2 \leq 2$

$x_1 + 3x_2 \geq 3$

$x_2 \leq 4$

$x_1, x_2 \geq 0$

17. Explain VAM method with example

18. Explain dual simplex method and find the dual of the following LPP

$$\text{Max } z = 50x_1 + 30x_2$$
$$\text{Subject to } 4x_1 + 3x_2 \leq 100$$
$$3x_1 + 5x_2 \leq 150$$
$$x_1, x_2 \geq 0$$

Course Code: SJCSDC08

Name of the Course: R Programming for Data Analytics

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)	Lab(Hrs)
CO1	Understand the basic programming structure of R– Data frame, Matrix, List, Packages and Functions	PSO2	U	P	5	5
CO2	Understand various visualization models and their inference – Scatter plots, histogram, boxplot	PSO2	A	P	5	5
CO3	Apply statistical functions, models and their Inferences – Central tendency measure, Range, Variance, Standard Deviation	PSO5	A	P	5	5
CO4	Use data normalization for domain specific dataset	PSO3	A	P	5	5
CO5	Apply distribution models, Regression models and ANOVA	PSO5	Z	P	5	5
CO6	Design data model, visualization and inference of dataset to gain insights	PSO2	Z, C	P	5	5

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : R PROGRAMMING FOR DATA ANALYTICS

Course Code :SJCSDC08

Course Objectives

To impart knowledge to make the students

1. To understand the basics constructs of R Programming and Visualization.
2. To understand and apply Exploring variables using Visualization.
3. To understand and apply Inferential Statistics and Regression Models.

UNIT I

Introduction:What is R–Downloading and Installing R–. **Getting Data into R:** First Step in R:Typing in Small Datasets – Concatenating Data with c Function – Combining Variables with the c, cbind, rbind Functions - Vector Function –Matrix - Ddata frame – List - Importing Excel Data – Accessing Data from other Statistical Packages – Accessing the Database. Functions - The Attach Function – Exporting Data - The Tapply Function – The Supply and Lapply Function – The Summary and Table Function.

UNIT II

Importing Data – Csv, Excel, Table, Xml, Json , Databases Conditional – Control flow – Loops – A Function with Multiple Arguments - Cleaning Data : – Exploring raw data –Missing values - Zeros and NAs – Separating– Uniting Columns - String Manipulation – Filling Missing values – Packages – R Visualization Packages – Lattice – ggplot2 – understanding plots – aesthetics - - statistical function - Histogram – Box Plot – Density Plot– Scatter Plots The Plot Function –Adding a Smoothing Line The Pie Chart – The Bar and Strip Chart – Box Plot – Cleveland Dotplots- Reporting– Data Preparation – Embedding R chunks – Labelling and reusing code chunks – Report Compiling – Configuring – R Packages – shiny - ggvis

UNIT III

Variable Analysis – One variable – Understanding outliers through – histogram , boxplot, density plot – dataset– pseudo dataset of facebook Exploring two variables – Understanding Variables and relationships – scatter plots – correlations – condition means – Explore multivariate variables – Visualization of variables using aesthetics in R – Case study – Explore Diamond dataset for prize prediciton

UNIT IV

Data types – Categorical – Binary – ordinal – Nominal – Continuous – Discrete – Data Dimensions – Univariate – bivariate – multivariate – Numerical Measures – Central Tendency – Mean – Median – Mode - Understanding data using central tendency – plotting histogram – density plots and inference of plot - Variability Measure – Variance - Range - IQC - and Standard Deviation – Sum of squares – Squared Deviations– Absolute Deviations - Identify outlier using Inter Quartile Range – Visualization using boxplot

UNIT V

Data standardizing – Z Score – Negative Z Score – Continuous Distributions - Compute proportions – Relative Frequency histogram - Normalized Distribution using Ztable – Probability Distributions -

Probability of mean – location of mean distribution - Sampling Distributions — Klout Sampling Distribution – Understanding Shape of Distribution – Standard Error - Standard Deviation of sampling distribution – Ratio of Sampling Distribution-Central Limit Theorem R – Mean of sample means Advanced Analytics Regression Analysis – Simple Regression Analysis - – Logisitc Regression – Multiple Regression ANNOVA Model

REFERENCE

1. Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters,"A Beginner"s Guide to R" Springer, 2009
2. Roger D. Peng, "R Programming for Data Science" Lean Publishing, 2014
3. R Data camp – Online Course Contents - https://campus.datacamp.com/courses/

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS
SEMESTER II
MODEL QUESTION PAPER
SJCSDC08 R PROGRAMMING

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

Each question carries ***2*** *weightage.*

1.
   a. **Explain what is R?**
   b. **List out some of the function that R provides?**
   c. In R how you can import Data?
   d. How can you save your data in R?

2.

a) **Explain what is With () and By () function in R is used for?**
b) **What are the data structures in R that is used to perform statistical analyses and create graphs?**
c) **In R how missing values are represented ?**
d) **Explain what is transpose?**
e) What is the use of subset () function and sample() function in R ?

3. Explain Apply functions?

4. Define open-source **R** package shiny

5.How to understanding the outliers through EDA?

6.Illustrate the Univariate, bivariate and multivariate

7. What is negative Z-SCORE

Part B

Answer **any four** *questions.*
Each question carries ***3*** *weightages.*

8. Explain datatypes with examples?
9. Explain the central tendency measures?
10. Explain data cleaning procedure in R programming?
11. What is R shiny and mention its features?
12. Explain logistic regression with example.
13. Explain standard error and standard deviation in detail.
14. Explain advanced analytics regression.

Part C

Answer **any two** *questions.*
Each question carries ***5*** *weightages*

15. Explain EDA in detail.
16. Explain the built-in functions in R Programming.
17. How do you detect outliers in EDA?
18. Explain ANOVA model in detail.

Course Code: SJCSDC09

Name of the Course: Data Visualization

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)
CO1	Understand the concepts of visualization	PSO1	U	C	10
CO2	Understand the methods for visualizing data in D3j, c3j, and Tableau	PSO1	R	P	10
CO3	Apply Visualization methods for different data domains	PSO2	Z	P	10
CO4	Design Interactive Charts based on Data	PSO5	A	P	10
CO5	Distinguish and Suggest the appropriate data visualization tools for domain specific applications	PSO1	Z	C	10
CO6	Design an Interactive data visualization story board for data	PSO5	E	P	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : DATA VISUALIZATION

Course Code :SJCSDC09

Course Objectives

To impart knowledge to make the students

1. To understand how accurately represent voluminous complex data set in web and from other data sources

2. To understand the methodologies used to visualize large data sets
3. To know how to work with visualization tools.

UNIT I

INTRODUCTION: Context of data visualization – Definition, Methodology, Visualization design objectives. Key Factors – Purpose, visualization function and tone, visualization design options – Data representation, Data Presentation, Seven stages of data visualization, widgets, data visualization tools. **VISUALIZING DATA METHODS:** Mapping - Time series - Connections and correlations - Scatter plot maps - Trees, Hierarchies and Recursion - Networks and Graphs, Info graphics.

UNIT II

INTERACTIVE DATA VISUALIZATION: Introduction to D3 **-** Fundamental Technology - Drawing with data – Scales – Axes – Updates, Transition and Motion – Interactivity - Layouts – Geomapping – Exporting-**Data to create Visualizationwith SVG** - SVG – Styling CSS – Shapes – SVG Properties – SVG Text - Drawing – Transformations – Building Chart with SVG (Scalable Vector Graphics) - Shaping Web Pages – Selections – Attributes – Chaining Methods –Data Joins - Sizing – scales – axes – Loading – Filtering – Interactive Charts – Buttons using Data Join – Transition using Key

UNIT III

D3-BASED REUSABLE CHART LIBRARY: Introduction to D3 – Setup and Deployment – Generate Chart– Customize Chart – How to Use APIs – Customize Style – Building Real time and Live Updating animated graphs with C3.

UNIT IV

TABLEAUE INTRODUCTION: Environment Setup – Navigation – File & Data Types. **DATA SOURCE:** Custom Data View – Extracting Data – Fields Operations – Editing Meta Data – Data Joining – Data Blending. Worksheets

UNIT V

TABLEAUE CHARTS: Bar Chart – Line Chart – Pie Chart – Scatter Plot – Bubble Chart –Gantt Chart –Histograms **-** Waterfall Charts. **ADVANCED:** Dashboard – Formatting – Forecasting – Trend Lines

REFERENCES

1. Ben Fry, "Visualizing Data", O"Reilly Media, Inc., 2007.
2. Scott Murray, "Interactive data visualization for the web", O"Reilly Media, Inc., 2013.
3. Ritchie S. King - *Visual Storytelling with D3 – An Introduction to Data Visualization with D3,* Addison-Wesley- *Data Analytic Series,* ISBN 10: 0321933176

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS
SEMESTER 1I
MODEL QUESTION PAPER
SJCSDC09 DATA VISUALIZATION

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightage.*

1.What all the key factors of visualization?

2.Explain about data visualization tools

3.Illustrate 7 stages of data visualization

4. Explain D3

5.Explain the fundamental technology of visualization

6.What is the role of SVG in visualization

7.How to generate charts through D3

Part B

Answer **any four** *questions.*

*Each question carries **3** weightages*

8.Describe data visualization methods

9.How visualization influenced time series

10.Steps for styling CSS

11.What is customize charts and customize style

12.Describe the environment setup for TABLEAUE

13.Purpose of TABLEAUE

14.Describe advanced visualization methods using Tableaue

Part C

Answer **any two** *questions.*

Each question carries ***5*** *weightages*

15.Explain Hierarchies and Recursion

16.Define SVG properties

17.What all are the data sources in Tableaue

18.Describe about Tableaue charts with diagrams

Course Code: SJCSDC10

Name of the Course: Linear Algebra for Machine Learning

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**
CO1	Understand the basics of Linear Programming constructs	PSO1	A	P	6
CO2	Apply vector spaces and their applications in Machine Learning	PSO5	A	P	6
CO3	Understand the concepts of matrix and Gaussian Elimination	PSO1	A	P	6
CO4	Understand the concepts of differential equations	PSO1	A	P	6
CO5	Apply the concepts of Linear Algebra in Machine Learning Algorithms	PSO5	E	C	6

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : LINEAR ALGEBRA FOR MACHINE LEARNING

Course Code :SJCSDC10

Course Objectives

To impart knowledge to make the students

1.To understand concepts basics concepts of Linear Algebra

2.To understand concepts of vector spaces and matrices

3.To understand the applications of Linear Algebra in Machine Learning

UNIT I

The Geometry of Linear Equations- An Example of Gaussian Elimination- Matrix Notation andMatrix Multiplication - Triangular Factors and Row Exchanges- Inverses and Transposes

UNIT II

Vector Spaces: Vector Spaces and Subspaces – Solving Ax=0 and Ax=b - Linear Independence, Basis and Dimension- The Four Fundamental Subspaces- Graphs and Networks- Linear Transformations

UNIT III

Determinants: Introduction- Properties of the Determinant- Formulas for the Determinant - Applications of Determinants

UNIT IV

Eigenvalues and Eigenvectors: Introduction- Diagonalization of a Matrix .- Difference Equations and Powers A^k- Differential Equations and e At - Complex Matrices- Similarity Transformations – A - Applications of Machine Learning – Use cases

UNIT V

Positive Definite Matrices: Minima, Maxima, and Saddle Points - Tests for Positive Definiteness-Singular Value Decomposition – Machine Learning Applications – Use cases

Text Book

1. Gilbert Strang(2006). Linear Algebra and Its Application, Fourth Edition, Academic Press.

Reference Books

1. David C. Lay, Steven R. Lay, Judi J. McDonald (2014). Linear Algebra and Its Applications, Pearson Education.

2. Peter D. Lax(2007). Linear Algebra and Its Applications, Second Edition, Wiley Publication

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS
SEMESTER II
MODEL QUESTION PAPER
SJCSDC10 LINEAR ALGEBRA FOR MACHINE LEARNING

TIME: 2 ½ HOURS **MAX:30WEIGHTAGE**

Part A

Answer **any four** *questions.*

Each question carries ***2*** *weightages.*

1. Solve the following equations by Gauss Elimination Method.

x+4y-z = -5

x+y-6z = -12

3x-y-z = 4

2. Prove that any two bases of finite dimensional vector space V have same number of elements.

3. Define Linear Transformation.

4.Explain the properties of determinants.

5. Prove that any set of vectors which contains the zero vector is linearly dependent

6. **a)** Prove that the column vectors of every 3×5 matrix AA are linearly dependent.
(b) Prove that the row vectors of every 5×3 matrix BB are linearly dependent.

7. What are the four fundamental subspaces?

Part B

Answer **any four** *questions.*

*Each question carries **3** weightages*

8. Mention the applications of determinants.

9. Let A be an n×n matrix with real number entries. Show that if A is diagonalizable by an orthogonal matrix, then A is a symmetric matrix.

10. Solve the following system of linear equations:

4y+z=2; 2x+6y-2z=3; 4x+8y-5z=4

11. What are Eigen value and Eigen vector? Explain its significant in big data analytics.

12. Explain positive definite and positive semi definite matrixes.

13. Define Matrix, Diagonal matrix, Symmetric matrix.

14. Define feasible solution and basic feasible solution.

Part C

Answer **any two** *questions.*

*Each question carries **5** weightages*

15. Define vector spaces. Show that the set V is a vector space over the field of reals.
16. Explain Singular Value Decomposition in detail.
17. Explain minima, maxima and Saddle points.
18. Solve the following system using Gauss Jordan Operation elimination method: x+2y-z=0; 2x+y-z=0; 7x+14y-21z=0

Course Code: SJCSDL02

Name of the Course: Lab2: R Programming and ADBMS

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Lab(Hrs)**
CO1	Configure R environment for development of application, Develop functional application in using r scripting	PSO1	A	P	5
CO2	Develop application which process CSV, XML, JSON, XML, Develop application with visualization - bar Chart, line graph, box	PSO1	A	P	5

	plot , Histogram and Scatter Plots.				
CO3	Apply basic statistical operation using R	PSO5	A	P	5
CO4	**Use** an SQL interface of a multi-user relational DBMS package to create, secure, populate, maintain, and query a database.	PSO4	A	P	5
CO5	**Formulate** query, using SQL, solutions to a broad range of query and data update problems	PSO5	A	P	5
CO6	**Transform** an information model into a relational database schema and to use a data definition language and/or utilities to implement the schema using a DBMS.	PSO4	A	P	5

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Part A- R programming

- **Exploratory Data Analytics**- Histogram and Box plot Analysis and Violin plot analysis
- Apply Functions
- Pre processing and Data wrangling
- Missing value Imputation
- DPLYR Package and its Functions
- Dashboard
- **Machine Learning Models-**K means clustering, Simple linear regression, Multi linear Regression, Decision Tree Classification
- Text Analytics

Part B- ADBMS

- **SQL QUERY**- DDL, DML, Aggregate Functions, Join Query, Views

- **PL/SQL**- Functions, Stored Procedure, Case study- MySQL
- **MONGO DB-** CRUD Operations, Array Query, Aggregate Functions, Working with CSV file, Case study- Mongo DB

SEMESTER III

Course Code: SJCSDC11

Name of the Course: Cloud Computing

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)
CO1	Understand Cloud Service Models and architectures	PSO1	U	C	5
CO2	Apply the concepts of Virtualization	PSO4	A	P	10
CO3	Understand cloud programming models	PSO1	A	P	10
CO4	Analyze various cloud services and features of cloud service providers	PSO4	Z	P	10
CO5	Critically analyze case studies to derive the cloud models for developing and deploying	PSO5	E	P	10

	cloud-based applications				
CO6	Understand the Risks and Management of Cloud environments	PSO1	U	C	5
CO7	Compare the cloud service models and providers to suit business needs	PSO2	Z	P	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : CLOUD COMPUTING

Course Code :SJCSDC11

Course Objectives

To impart knowledge to make the students

1. To understand cloud service models, architecture, programming model, security and familiarize with leading service cloud providers to critically evaluate the cloud services for business applications of various domains

UNIT I

Technologies for Network-Based System – System Models for Distributed and Cloud Computing – NIST Cloud Computing Reference Architecture. Cloud Models:- Characteristics – Cloud Services – Cloud models (IaaS, PaaS, SaaS) – Public vs Private Cloud –- Cloud ecosystem – Service management – Computing on demand.

UNIT II

VIRTUALIZATION Basics of Virtualization - Types of Virtualization - Implementation Levels of Virtualization - Virtualization Structures - Tools and Mechanisms -Virtualization of CPU, Memory, I/O Devices - Virtual Clusters and Resource management – Architectural Design of Compute and Storage Clouds – Layered Cloud Architecture Development – Design Challenges - Inter Cloud Resource Management

UNIT III

Parallel and Distributed Programming Paradigms – MapReduce , Twister and Iterative MapReduce – Hadoop Library from Apache – Mapping Applications - Cloud Software Environments -Eucalyptus, Open Nebula, OpenStack,

UNIT IV

Cloud Computing Providers (Usecase on any one Service Provider) - Infrastructure as a Service (IaaS) providers - Google Compute Engine, Amazon AWS, Microsoft Azure; Cloud Database providers - Google Cloud SQL, Microsoft Cloud SQL Database; Cloud Storage providers: Google Drive API, Google Cloud Storage, Azure Blob Storage.Platform as a Service (PaaS) providers for Web Rapid Application Development (RAD) - Google App Engine; Distributed Storage providers: Google Cloud Datastore, Azure tables; Distributed Computing providers and frameworks: Google Cloud Dataflow, Apache Spark –- Cloud Economics : Cloud Computing infrastructures - Cloud platform for an organization, -economic constraints and business needs (e.g Amazon, Microsoft and Google, Salesforce.com, Ubuntu and Redhat) Domain Specific –Analysis of Case Studies - adopt cloud computing architecture

UNIT V

Cloud Security Challenges and Risks – Software-as-a-Service Security – Security Governance – Risk Management – Security Monitoring – Security Architecture Design – Data Security – Application Security – Virtual Machine Security - Identity Management and Access Control – Autonomic Security.

REFERENCES

1. Thomas Erl, "Cloud Computing: Concepts, Technology & Architecture", Prentice Hall, May 2013
2. Kai Hwang, Geoffrey C Fox, Jack G Dongarra, “Distributed and Cloud Computing, From Parallel
   Processing to the Internet of Things”, Morgan Kaufmann Publishers, 2012.
3. John W.Rittinghouse and James F.Ransome, “Cloud Computing: Implementation, Management, and Security”, CRC Press, 2010.
4. GautamShroff, “Enterprise Cloud Computing Technology Architecture Applications”, Cambridge University Press; 1 edition, [ISBN: 978- 0521137355], 2010.
5. Toby Velte, Anthony Velte, Robert Elsenpeter, “Cloud Computing, A Practical Approach”, TMH, 2009.
6. Kumar Saurabh, “ Cloud Computing – insights into New-Era Infrastructure”, Wiley India,2011 George Reese, “Cloud Application Architectures: Building Applications and Infrastructure in the Cloud” O'Reilly
7. James E. Smith, Ravi Nair, “Virtual Machines: Versatile Platforms for Systems and Processes”,Elsevier/Morgan Kaufmann, 2005.
8. Katarina Stanoevska-Slabeva, Thomas Wozniak, SantiRistol, “Grid and Cloud Computing – A Business Perspective on Technology and Applications”, Springer.

Model Question Paper

ST. JOSEPH’S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS - SEMESTER 3 MODEL QUESTION

PAPER
SJCSDC11 CLOUD COMPUTING

TIME: 2 ½ HOURS **MAX: 30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightage.*

1. What do you mean by Cloud eco system?
2. What is Open Nebula?
3. Define Data Security.
4. What do you mean by Virtual machine security?
5. Elaborate the NIST reference architecture of cloud computing with diagram.
6. Describe cloud computing models in detail.
7. How to implement the virtualization in various level? Explain with example.

Part B

Answer any **four**questions
Each question carries a weightage of 3

8. Elaborate the usage of Google drive with example.
9. Explain the various economic constraints of cloud in detail.
10. Explain Software as a Service Security in detail.
11. Elucidate the Identity management and access control in detail.
12. Explain layered cloud architecture development in detail.
13. Discuss the mapping applications in detail.
14. Write short note on the following
    - Eucalyptus
    - OpenStack

Part C

*Answer any **two** questions*

Each question carries a weightage of 5

15. How to adopt cloud architecture for an organization? Explain with example.
16. Analyze the role and need of cloud storage providers for running an E-Commerce application for a business.
17. Explain Cloud security challenges and risks
18. Elucidate about Cloud platform for an organization

Course Code: SJCSDC12

Name of the Course: Map Reduce Programming

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**	**Lab(Hrs)**
CO1	Understand MapReduce Processing architectures	PSO1	U	C	4	0
CO2	Configure and setup MapReduce Processing architectures Ecosystem – Hadoop, Spark, Pig and Hive	PSO6	A	P	4	5
CO3	Understand and write MapReduce program using Pig and Hive, spark	PSO2	A	P	4	5
CO4	Analyze dataset using Pig and Hive	PSO3	Z	P	5	5

CO5	Critically analyze case studies for and suggest MapReduce Programming models based on domains and applications	PSO5	E	P	4	5
CO6	Distinguish Hadoop and SPARK	PSO2	Z	P	4	5
CO7	Design and setup a Big Data Analytics Ecosystem for specific Business scenarios	PSO5	E	P	5	5

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : MAP REDUCE PROGRAMMING

Course Code :SJCSDC12

Course Objectives

To impart knowledge to make the students

1. To understand MapReduce programming architecture, processing models.
2. To understand and design MapReduce Programming using PIG and Hive
3. To understand and compare the architectural and processing of MapReduce Programing languages Pig, Hive and SPARK

UNIT I

Introduction to Big Data – Distributed file system –, Map Reduce Algorithm- Hadoop Storage [HDFS], Common Hadoop Shell commands - Anatomy of File Write and Read, NameNode, Secondary NameNode, and DataNode - Hadoop Configuration – SPARK Configuration - Pig Configuration – Hive Configuration

UNIT II

Pig Introduction : Overview of Pig - Pig Architecture - Pig Execution modes, Pig Grunt shell and Shell - commands. Pig Latin Basis: Data model, Data Types, Operator - Pig Latin Commands - Load & Store, Diagnostic Operators, Grouping, Cogroup, Joining, Filtering, Sorting, Splitting - Built-In Functions,

User define functions.- Pig Execution Modes – Batch Mode – Embedded Mode – Pig Execution in Batch Mode – Embedding Pig in Python – Use cases - Map Reduce programs with Pig – Pig Vs SQL

Unit III

Introduction of Hive - Hive Features - Hive architecture -Hive Meta store - Hive data types – Hive Tables - Table types - Creating database , Altering database, Create table, alter table, Drop table, - Built-In Functions - Built-In Operators, User defined functions, - View – Pig Vs Hive

UNIT IV

HiveQL–Introduction to HiveQL, HiveQL Select, HiveQL – MapReduce using HiveQLOrderBy,Group By Joins, LIMIT, Distribute By , Cluster By - Sorting And Aggregation – Partitioning – Static –Dynamic – Index Creation - Bucketing – Analysis of MapReduce execution – Hive Optimization – Setting Hiivng Parameters. – Usecase :MapReduce using Hive QL – HiveQLVs SQL

UNIT V

SPARK – MapReduce - RDD Transformations – SPARK Operations – Usecase with SPARK and Comparison-MapReduce – Python – R – Pig – Spark – Hadoop - Limitations – Advantage – SPARK vsHadoop – SPARK Vs Pig and Hive – MapReduce

REFERENCES

1. Boris Lublinsky Kevin T. Smith Alexey Yakubovich ,ProfessionalHadoop® Solutions, Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk deroos et al. , “Understanding Big data ”, McGraw Hill, 2012.

Tom White ,“Hadoop: The Definitive Guide”, O'Reilly Media 3rd Edition,May6, 2012

Model Question Paper

ST. JOSEPH’S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS - SEMESTER 3 MODEL QUESTION PAPER

SJCSDC12 MAP REDUCE PROGRAMMING

TIME: 2 ½ HOURS **MAX: 30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightage.*

1. List and brief on the pig execution modes.
2. What is Hive Meta Store?
3. How does index function in HiveQL?
4. Where is 'R' Language used?
5. Demonstrate any 10 Shell commands in Hadoop.
6. Explain the File Write and Read operations in Hadoop.
7. Write a detailed note on Pig's Data Model (data types)

Part B

Answer any **four**questions

Each question carries a weightage of 3

8. Give the Categories of Built-in functions and brief any category in detail.
9. Illustrate the Hive create, alter and drop table commands with example.
10. Describe Views in HiveQL.
11. Elaborate with examples the various types of joins in HiveQL.

12. Compare HiveQL with SQL.
13. Discuss about RDD transformations.
14. Differentiate:
    - SPARK and Hadoop
    - SPARK Vs. Pig and Hive

Part C

*Answer any **two** questions*

Each question carries a weightage of 5

15. Elaborate the HDFS Architecture and its elements.
16. Draw and explain the MapReduce execution.
17. Explain RDD Transformation
18. Illustrate a usecase : MapReduce using Hive QL

Course Code: SJCSDC13

Name of the Course: Machine Learning

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**	**Lab(Hrs)**
CO1	Understand the concepts of machine learning	PSO1	U	C	5	0
CO2	Understand the theoretical concepts of probabilistic and linear methods	PSO1	R	C	5	0
CO3	Distinguish Supervised, Unsupervised and semi supervised learning	PSO2	Z	P	5	6
CO4	Understand and Apply the algorithms for a given specific problem in a specific tool	PSO3	A	P	5	8
CO5	Suggest Supervised, Unsupervised and semi supervised algorithms for specific application	PSO3	Z	P	5	8

CO6	Design a Machine Learning system for any specific domain	PSO5	E	P	5	8

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : MACHINE LEARNING

Course Code :SJCSDC13

Course Objectives

To impart knowledge to make the students

1.To understand the concepts of machine learning concepts, algorithms, and probabilistic, lineargraphical models.

2.To apply the machine learning algorithms for various applications.

UNIT I

Introduction : Machine Learning - Machine Learning Foundations –Overview – applications - Types of machine learning - basic concepts in machine learning Examples of Machine Learning -Applications - Linear Models for Regression - Linear Basis Function Models - The Bias-Variance Decomposition - Bayesian Linear Regression - Bayesian Model Comparison

UNIT II

Supervised Learning Linear Models for Classification - Discriminant Functions -Probabilistic Generative Models - Probabilistic Discriminative Models - Bayesian Logistic Regression. Decision Trees - Classification Trees- Regression Trees - Pruning. Neural Networks -Feed-forward Network Functions - Error Backpropagation - Regularization - Mixture Density and Bayesian Neural Networks - Kernel Methods - Dual Representations - Radial Basis Function Networks. Support Vector Machines - Ensemble methods- Bagging-Boosting – Evaluation Methods

UNIT III

Unsupervised Learning Clustering- K-means - EM - Mixtures of Gaussians - The EM Algorithm in General - Model selection for latent variable models - high-dimensional spaces -- The Curse of Dimensionality - Dimensionality Reduction - Factor analysis - Principal Component Analysis - Probabilistic PCA- Independent components analysis

UNIT IV

Probabilistic Graphical Models Directed Graphical Models - Bayesian Networks - Exploiting Independence Properties - From Distributions to Graphs -Examples -Markov Random Fields - Inference in Graphical Models - Learning –Naive Bayes classifiers-Markov Models – Hidden Markov Models – decoding states from observations, learning HMM parameters-Inference – Learning Generalization –

Undirected graphical models-Markov random fields- Conditional independence properties - Parameterization of MRFs - Examples - Learning - Conditional random fields (CRFs) - Structural SVMs

UNIT V

Advanced Learning Sampling – Basic sampling methods – Monte Carlo. Reinforcement Learning- K-Armed BanditElements - Model-Based Learning- Value Iteration- Policy Iteration. Temporal Difference Learning Exploration Strategies- Deterministic and Non-deterministic Rewards and Actions- Eligibility Traces - Generalization- Partially Observable States- The Setting- Example. Semi - Supervised Learning. Computational Learning Theory - Mistake bound analysis, sample complexity analysis,

REFERENCES

1. Christopher Bishop, "Pattern Recognition and Machine Learning" Springer, 2006
2. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012

3. EthemAlpaydin, ―Introduction to Machine Learning 3e (Adaptive Computation and Machine Learning Series)‖, Third Edition, MIT Press, 2014

4. Tom M Mitchell, ―Machine Learning‖, First Edition, McGraw Hill Education, 2013.

5. Hastie, Tibshirani, Friedman, "The Elements of Statistical Learning" (2nd ed)., Springer, 2008

6. Stephen Marsland, "Machine Learning –An Algorithmic Perspective", CRC Press, 2009

7. www.swayam.com

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS - SEMESTER 3 MODEL QUESTION PAPER
SJCSDC13 MACHINE LEARNING

TIME: 2 ½ HOUR **MAX: 30WEIGHTAGE**

Part A

Answer **any four** *questions.*

Each question carries ***2*** *weightage.*

1. List out some popular machine learning algorithms?
2. Distinguish between classification and regression.
3. Write down the limitations of K-means clustering.
4. Define structured SVM.
5. Mention the applications of Reinforcement learning.
6. Illustrate Bayesian Model Comparison
7. Explain Probabilistic Principal Component Analysis

Part B

Answer any **four**questions

Each question carries a weightage of 3

8. Explain the Bias-Variance Tradeoff.
9. Consider a data set consisting of variables having more than 30% missing values? For example, out of 50 variables, 8 variables have missing values higher than 30%. How will you deal with them?
10. Consider a cancer detection data set List out the steps to be performed to build a classification model with 95% accuracy.
11. How can you choose a classifier based on training set size?
12. Explain Latent Dirichlet Allocation (LDA)
13. How is KNN different from K-means clustering?
14. Explain about Markov model with an example.

Part C

*Answer any **two** questions*

Each question carries a weightage of 5

15. How will you define the number of clusters in a K-means clustering algorithm?
16.
    - What is Back Propagation and Explain it's Working.
    - What is the role of Activation function?
17. A doctor knows that Cold causes fever 50% of the time – Prior probability of any patient having cold is 1/50,000 – Prior probability of any patient having fever is 1/20. If a patient has fever, what's the probability he has cold? (5)
18. Analyzetic-tac-toe moves by applying reinforcement learning

ELECTIVES

Course Code: SJCSDE01

Name of the course: Internet of Things

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)	Lab(Hrs)
CO1	To understand the basics of mental ability, logical ability through active involvement	PSO1	U	C	6	0
CO2	To understand analytical reasoning and apply in software applications	PSO2	U	C	6	0
CO3	To development of programming and database management skills	PSO6	A	P	6	10
CO4	To improve Presentation Skills, Controlling Nervousness and Stage Fright, Communication via Group Discussions	PSO6	A	P	6	10
CO5	To practice technical programming, cracking code, simple logic and concepts	PSO3	R, Z	P	6	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title: Internet of Things

Course Code : SJCSDE01

Course Objectives

To impart knowledge to make the students

1. To understand the concepts of micro controller
2. To apply IoT Applications for specific domains

3. To understand the Programming Fundamentals with C using Arduino

UNIT I

Introduction and definition to IOT - What is an IOT? - Explore the scenario for application of IOT Communication definitions Concepts - Capturing and Storing the data - What to do with the data…applying Expert Systems and Machine Learning; IOT Detailed understanding of Solution Architecture - IOT Device Architecture - IOT Network/Communication Architecture with an understanding on client server and looselycouple storage servers and message queues - IOT Application Architecture.

UNIT II

Programming Fundamentals with C using Arduino IDE - Understanding the Arduino IDE - Installing and Setting up the Arduino IDE - Connecting the Arduino IDE with devices - Program Structure in C - Basic Syntax-Data Types / Variables / Constants - Operators, Conditional Statements and Loops -Functions , Array and Pointers - Strings and I/O -Using Arduino C Library functions for Serial, delay and other invoking functions - Working with LED and Switch example on Arduino C Library functions

UNIT III

Working with Arduino for data acquisition with IOT Devices - Understanding Sensors and Devices - Understanding basic electronic components and power elements - Understanding the Inputs from Sensors - Working with Temperature Sensors -Working with Ultrasound Sensor -Working with humidity sensor - Working with Motion Sensor - Working with IR Sensor - Working with Proximity Sensor - Working with Photo Diode - Working with Accelerometer and vibration sensor - Measuring Voltage and Current

UNIT IV

Working with Arduino for data acquisition with IOT Devices - Understanding the Outputs - Activating LED Lights - Activating Relays - Activating Buzzer - Running DC Motors - Running - Stepper Motors and Servo Motors

UNIT V

Programming Fundamentals with Web Applications for handling Data Communication from IOT Device - Understanding the data capture through web services - Creating and Programming a rest web service with ASP.NET / PHP - Calling and accessing the Web Service in a Client; Building and Using Communication Devices to data transfer from IOT Devices - Understanding the Communication Principles to Transfer the data from IOT Devices; Remote Communication to cloud/external application - Using WIFI to Transfer the data from IOT Sensor.

REFERENCES

1. Michael Margolis, “Arduino Cookbook” 2nd Edition, O'Reilly Media, 2011

2. Michael Collier, Robin Shahan, “Fundamentals of Azure”, Microsoft Press, 2015, ISBN: 978-0-7356-9722-5

3. Rick Rainey, “Azure Web Apps for Developers”, Microsoft Press, 2015, ISBN: 978-1-5093-0059-4

4. Microsoft Azure, “Introduction to Microsoft Azure Storage”, https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction

Model Question Paper

Course Code: SJCSDE02

Name of the Course: Text Analytics

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)	Lab(Hrs)
CO1	Understand the basic issues of and types of text mining	PSO1	U	C	5	0
CO2	Appreciate the different aspects of text categorization and clustering	PSO4	A	P	5	6
CO3	Understand the role played by text mining in information retrieval extraction	PSO2	U	C	5	0
CO4	Apply the probabilistic models, clustering and classification for text mining	PSO5	A	P	5	8
CO5	Analyze the current trends in text mining	PSO4	Z	P	5	8
CO6	Design a text analytic framework to analyze text data for specific domain	PSO5	A	P	5	8

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : TEXT ANALYTICS

Course Code :SJCSDE02

Course Objectives

To impart knowledge to make the students

1. To understand the knowledge of text mining and pre-processing techniques
2. To understand and apply the data mining classification techniques
3. To understand and apply Probabilistic Model for text mining
4. To apply the text mining approaches with case studies

UNIT I :

TEXT MINING

Text Mining - Definition - General Architecture – Core Text mining Operations. NLP, nature of unstructured and semi-structured text. **Pre-processing**- Textual information to numerical vectors - Collecting documents - document standardization and Representation - tokenization - lemmatization - stemming - Parsing text - keywords, n-grams - POS, Corpus - sentence boundary determination - vector generation for prediction

UNIT II:

TEXT CATEGORIZATION

Text Categorization – Definition – knowledge engineering, Machine Learning- **Classification**-Decision Tree Classifiers -Rule- based Classifiers - Probabilistic and Naive Bayes Classifiers - Linear Classifiers- Classification of Linked and Web Data - Meta-Algorithms– **Clustering** – Definition- Distance - based Algorithms-Word and Phrase-based Clustering -**Semi- Supervised** Clustering -Transfer Learning

UNIT III:

INFORMATION RETRIEVAL AND INFORMATION EXTRACTION

Information retrieval- keyword search - Vector space scoring, Models - web- based document search-matching-inverted lists-evaluation. **Information extraction**-Architecture - Co-reference - Named Entity and Relation Extraction-Template filling and database construction –Applications. Inductive -Unsupervised Algorithms for Information Extraction. **Text Summarization Techniques** -Topic Representation -Influence of Context -Indicator representations - Pattern Extraction - Apriori Algorithm – FP Tree algorithm - **Results summaries**

UNIT IV:

PROBABILISTIC MODELS

Probabilistic Models for Text Mining -Mixture Models -Stochastic Processes in Bayesian Nonparametric Models - Relationship Between Clustering, Dimension Reduction and Topic Modeling - Interpretation and Evaluation -Probabilistic Document Clustering and Topic Models - **Probabilistic Models for Information Extraction** -Hidden Markov Models -Stochastic Context-Free Grammars - Maximal Entropy Modeling -Maximal Entropy Markov Models - Conditional Random Fields

UNIT V:

VISUALIZATION METHODS AND TEXT MINING APPLICATIONS

Visualization Approaches -Architectural Considerations –Common Visualization Approaches for text mining - Applications - Mining Text Streams -Text Analytics in Social Media - Opinion Mining and Sentiment Analysis -Document Classification - Opinion Spam Detection - Case studies - Email filtering, Assigning Topics to news articles, extracting named entities,etc.,

REFERENCES

1. Sholom Weiss,Nitin Indurkhya, Tong Zhang,Fred Damerau"The Text MiningHandbook:Advanced Approaches in Analyzing Unstructured Data,", Springer, paperback2010

2. Ronen Feldman, James Sanger-"The Text Mining Handbook: Advanced Approachesin Analyzing Unstructured Data"-Cambridge University press,2006.

3. Charu C. Aggarwal ,ChengXiangZhai,Mining Text Data, Springer; 2012.

4. Introduction to Information Retrieval. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schuetze, Cambridge University Press, 2007.

5. Markus Hofmann, Andrew Chisholm "Text Mining and Visualization: Case Studies Using Open-Source Tools,", CRC press, Taylor & Francis,2016

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS - SEMESTER 3 MODEL QUESTION PAPER

SJCSDE03 TEXT ANALYTICS

TIME: 2 ½ HOUR **MAX: 30WEIGHTAGE**

Part A

Answer **any four** *questions.*

*Each question carries **2** weightages.*

1. Define general architecture of text mining.
2. What all are the preprocessing steps in text mining?
3. What is text categorization and define it?
4. Explain the role of word and phrase-based clustering in text mining.
5. Briefly explain the importance of Information extraction in text mining?
6. Explain text mining according to probabilistic model.
7. Explain common visualization approaches for text mining.

Part B

*Answer any **four** questions*

Each question carries weightage of 3

8. Give the advantages of text mining and its operations.
9. Write any classification method in text categorization.
10. Illustrate Apriori algorithm with example.
11. What is the purpose of unsupervised algorithm for information extraction?
12. Explain any two probabilistic models for information extraction
13. What is opinion spam detection?
14. Mention the applications of text analytics in social media with help of visualization.

Part C

*Answer any **two** questions*

Each question carries a weightage of 5

15. Explain preprocessing methods on the basis of text analytics about the domain E-mail filtering.
16. Explain the data sources for text analytics.
17. What are the methods of text categorization?
18. What is the relation between text analytics and sentiment analysis?

Course Code: SJCSDE03

Name of the course: Social Media Mining

	Course Outcome	POs / PSOs	CL	KC	Class Sessions (appr.)
CO1	Understand the concepts of Graph Models, social communities	PSO1	U	C	10
CO2	Understand the network models and measures to evaluate information	PSO1	R	C	10
CO3	Understand and apply algorithms to model data using graph and network structures and recommendations	PSO2	Z	P	10
CO4	Brief on algorithms on social data diffusion and apply for various domains	PSO4	A	P	10
CO5	Distinguish and Suggest the appropriate algorithms for domain specific applications for data modeling and information diffusion	PSO2	Z	P	10
CO6	Evaluate the algorithms for metrics	PSO4	E	P	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : SOCIAL MEDIA MINING

Course Code :SJCSDE02

Course Objectives

To impart knowledge to make the students

1. To understand how accurately analyze voluminous complex data set in socimedia and other sources
2. To understand the models and algorithms to process large data sets
3. To understand social behaviour and recommendation challenges and methodologies

UNIT I

Social Media Mining - Introduction – Atoms – Molecules – Interactions – Social Media mining Challenges - Graphs - Basics – Nodes – Edges – Degree of Distribution- Types –Directed – Undirected – Weighted - Graph Connectivity - Tress and Forests – Bipartite graphs – Complete Graphs – Sub graphs – Planar Graphs - Graph Representation - Graph Traversal Algorithms – Shortest path algorithms Dijkstra‟s - Spanning tree algorithms – Prims - Bipartite matching - Ford-Fulkerson algorithm

UNIT II

Network Models – Measures – Node : Eigen Centrality – Page Rank – Group Measures – Betweenness centrality - group degree centrality, centrality, and group - Closeness centrality - Node Linking Behavior - Transitivity and reciprocity - Linking Analysis - Cluster coefficient – Jaccard - Case Study : -Modeling small networks with real world model

UNITIII

Social media Communities – Social Communities – Member based Detection – Node degree – Node Similarity – Node reachability - Group Based detection methods - balanced – robust - modular – dense - hierarchical - Spectral Clustering : Balanced Community algorithm Community Evolution - Evaluation.

UNITIV

Social Network – Information Diffusion – Types - herd behavior - information cascades diffusion of innovation – epidemics – Diffusion Models Case Study – Herd Behavior – Information Cascades Methods – Social Similarity – assortativity – Social Forces - Influence homophily – Confounding - Assortativity measures – Influence measures – Predictive Models

UNITV

Recommendation Vs Search – Recommendation Challenges – Recommender algorithms - Content-Based Methods- Collaborative Filtering – Memory Based – Model Based – Social Media Recommendation – User friendship – Recommendation Evaluation – Precision – Recall Behavioral– User Behavior – User – Community behavior – User Entity behavior – Behavioral Analytics - Methodology

REFERENCES

1. Social Media Mining: An Introduction – Reza Zafarani , MohhammadAbiElasi – Published by Cambridge press, 2014 – (Free Ebook available http://dmml.asu.edu/smm/chapter**)**
2. Data Mining for Social Network Data- Memon, N., Xu, J.J., Hicks, D.L., Chen, H. (Eds.), Springer – Annals of Information Systems ,ISBN 978-1-4419-6287-4

Model Question Paper

Course Code: SJCSDE04

Name of the Course: Sentiment Analysis

	Course Outcome	**POs / PSOs**	**CL**	**KC**	**Class Sessions (appr.)**
CO1	The sources of user conversation, document types, identification of human emotions and design the methods to extract and classifying the content	PSO1	U	C	10
CO2	Design and prepare the sources of sentiment analysis problem, with expected input and output specification	PSO3	R, Z	C	10
CO3	Prepare and Apply the program using supervised and unsupervised learning, using a OSS tool	PSO5	A, E	P	10
CO4	Evaluate the opinion summarization and lexicon generation	PSO4	E	P	10
CO5	Design and prepare the sources of opinion summarization problem, with expected input and output specification	PSO4	R, Z	P	10
CO6	Prepare and Apply the code for the opinion summarization using algorithms in a OS tool.	PSO5	A, E	P	10

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Syllabus

Course Title : SENTIMENT ANALYSIS

Course Code :SJCSDE04

Course Objectives

To impart knowledge to make the students

1. Main objective of this course is to understand and practice on sentiment analysis from the source material (usually from text), by classifying the user conversation and extracts the opinions of subjective information using advanced tools and algorithms.

UNIT I

Sentiment Analysis Applications: Sentiment Analysis Research - Opinion Spam Detection. The Problem of Sentiment Analysis: Problem Definitions - Opinion Summarization - Different Types of Opinions - Subjectivity and Emotion - Author and Reader Standing Point. Document Sentiment Classification: Sentiment Classification Using Supervised Learning - Sentiment Classification Using Unsupervised Learning - Sentiment Rating Prediction - Cross-Domain Sentiment Classification - Cross-Language Sentiment Classification.

UNIT II

Sentence Subjectivity and Sentiment Classification: Subjectivity Classification - Sentence Sentiment Classification - Dealing with Conditional Sentences - Dealing with Sarcastic Sentences - Cross-language Subjectivity and Sentiment Classification - Using Discourse Information for Sentiment Classification.

UNIT III

Aspect-based Sentiment Analysis: Aspect Sentiment Classification - Basic Rules of Opinions and Compositional Semantics - Aspect Extraction - Identifying Resource Usage Aspect - Simultaneous Opinion Lexicon Expansion and Aspect. Extraction: Grouping Aspects into Categories - Entity, Opinion Holder and Time Extraction - Co reference Resolution and Word Sense Disambiguation.

UNIT IV

Sentiment Lexicon Generation: Dictionary-based Approach - Corpus-based Approach -

Desirable and Undesirable Facts. Opinion Summarization: Aspect-based Opinion Summarization - Improvements to Aspect-based Opinion Summarization - Contrastive View Summarization - Traditional Summarization Analysis of Comparative Opinions: Problem Definitions - Identify Comparative Sentences - Identifying Preferred Entities .

UNIT V

Opinion Search and Retrieval: Web Search vs. Opinion Search - Existing Opinion RetrievalTechniques Opinion Spam Detection: Types of Spam and Spamming - Supervised Spam Detection - Unsupervised Spam Detection - Group Spam Detection.

REFERENCES

1. Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies), Bing Liu, Morgan & Claypool Publishers (2012)

2. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Bing Liu, Cambridge University Press (2015)

3. http://nptel.ac.in/courses/106105158/61

4. Sentiment Analysis: Second Edition, GerardusBlokdyk ,Createspace Independent Publishing Platform (2018)

Model Question Paper

ST. JOSEPH'S COLLEGE (AUTONOMOUS) IRINJALAKUDA

M.Sc. DATA ANALYTICS - SEMESTER 3 MODEL QUESTION PAPER
SJCSDE04 SENTIMENT ANALYSIS

TIME: 2 ½ HOURS **MAX: 30WEIGHTAGE**

Part A

Answer **any four** *questions.*

Each question carries ***2*** *weightage.*

1. Define: Aspect-based Opinion Summary.
2. How to calculate the semantic orientation of the given phrase?

3. Define: Sarcastic Sentences.

4. What is the use of sentences level classification?

5. Describe about sentiment rating prediction. Specify the merits and demerits of Sentiment analysis for product rating..

6. Discuss the importance of cross domain and cross language sentiment classification.

7. What is Gradable comparison? Write its types.

Part B

*Answer any **four** questions*

Each question carries a weightage of 3

8. Discuss different types of sentence level sentiment classification with example.

9. Write the importance of Sentence level sentiment classification.

10. Discuss about the basic rules of opinions and compositional semantics.

11. Explain about four main approaches used in explicit aspect extraction.

12. Specify the use of Corpus-based approach in Sentiment Lexicon Generation.

13. Specify the improvements and refinements have been proposed by researchers for the basic aspect-based summary.

14. Elaborate with example about the types of Spam and Spamming.

Part C

*Answer any **two** questions*

Each question carries a weightage of 5

15. Discuss about Unsupervised Spam Detection.

16. How to prepare movie review data for sentiment analysis with step by step procedure?

17. Explain opinion summarization with example. Analyze opinion summarization techniques available.

18. Explain a scenario based on Sentiment Analysis

Course Code: SJCSDE05 /SJCSE06

Name of the course: Online Courses

	Course Outcome	POs / PSOs	CL	KC	Lab(Hrs)
CO1	Engage students at a deeper level.	PSO5	A	P	5
CO2	Improve digital literacy	PSO5	A	P	5
CO3	Exploring the subject knowledge and gets to implement.	PSO5	A	P	5
CO4	It gives real world skills, connects to the global world.	PSO5	A	P	5
CO5	It promotes life-long learning, the ability to focus on areas of improvement, options for flexible learning	PSO5	A	P	5

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Course Code: SJCSDL03

Name of the Course: Lab3: Map Reducing and Machine Learning

	Course Outcome	POs / PSOs	CL	KC	Lab(Hrs)

CO1	Implement a learning algorithm in Python, Predict classification or regression outcomes, with scikit-learn models in Python.	PSO1	A	P	5
CO2	Solve Non-linear problems using SVM. Apply machine learning algorithms to solve problems of moderate complexity.	PSO1	A	P	5
CO3	Apply and analyze dataset with machine learning algorithms	PSO5	A	P	5
CO4	Apply HIVE, Pig commands ,data analysis and word count using PIG	PSO4	A	P	5
CO5	Spark RDD Commands Data Analysis using Spark PySpark Commands Spark SQL commands	PSO5	A	P	5
CO6	Word count using Spark Word count using Map Reduce	PSO4	A	P	5

*R-Remember, U-Understand, A-Apply, Z-Analyze, E-Evaluate, C-Create

*F-Factual, C-Conceptual, P-Practical/Procedural

Part A- Map Reducing

- HIVE- DDL Commands
- HIVE- DML Commands
- HIVE- Joins and Indexing
- HIVE- View

- HIVE- Partitioning and Bucketing
- Pig Latin Commands
- Data Analysis using Pig
- Word Count using Pig
- Spark RDD Commands
- Data Analysis using Spark
- PySpark Commands
- Spark SQL commands
- Word count using Spark
- Word count using MapReduce

Part B- Machine Learning

- Regression model and Ridge regression using datasets.
- Multi linear regression using dataset.
- K means clustering on a dataset
- Hierarchical clustering on dataset
- Classification using decision tree on TITANIC dataset.
- Support Vector Machine (SVM) on a dataset
- Principal component Analysis (PCA) and K-means on a dataset
- Sentiment analysis on Twitter tweets.
- Web scraping using scrapy
- Bayesian Classifier to analysis on a dataset