Sonopant Dandekar Shikshan Mandali's
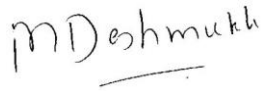
SONOPANT DANDEKAR ARTS, V.S. APTE COMMERCE AND M.H. MEHTA SCIENCE COLLEGE, Palghar.
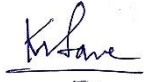
Programme	Certification Course on Machine Learning for Data Science using R
Objectives	R for data science focuses on the language's statistical and graphical uses. When students learn R for data science, they will learn how to use the language to perform statistical analyses and develop data visualizations. R's statistical functions also make it easy to clean, import and analyse data.
Duration of Training	40 sessions (Each session of 1 hour
Syllabus Committee	1)Prof.Rashmi Varade(course coordinator) 2)Dr.Yugandhara More 3)Dr.Janhavi Raut 4)Prof.Poornima Rane(N.b.Mehta College,Bordi)
Course Fee	Rs.300
Eligibility	FY/SY/TY (SCIENCE STUDENTS ONLY)
Examination	Online test and projects
Certificate	Certificate will be given to students who will pass examination.
Advisory Committee	1) Dr.Kiran Save (Principal) 2) Prof. Mahesh Deshmukh (IQAC Coordinator) 3) Prof.Rashmi Varade(course coordinator) 4) Appointed faculties
Course Content	**R- Installation [1 Hr]** R can be downloaded and installed on Windows, MAC OSX and Linux platforms from the Comprehensive R Archive Network (CRAN) webpage (http://cran.r-project.org/). After installing R software, install also the RStudio software available at: http://www.rstudio.com/products/RStudio/.
➢ **Introduction [2 Hrs]**	**Introduction [2 Hrs]** What is Data Science,Real-life examples and Applications, Data Scientist roles,

	Machine Learning vs. Data Science vs. AI, Machine Learning types, Generics of ML approaches.
	R programming Basics [4 Hrs] Basic arithmetic operations Basic arithmetic functions Assigning values to variables Basic data types
	Probability and Statistics for Data Science [4hrs] Basic probability theory, Random variables, Probability distributions, Markov models, Bayesian learning, Applications
	Regression Analysis [4hrs] Univariate linear regression, Multivariate linear regression, Polynomial Regression, Applications
	Classification[4Hrs] SVM, Multi-class SVM, Decision trees, K-NN, Applications.
	Clustering[4Hrs] Different clustering approaches and applications.
	Dimensionality Reduction [3 Hrs] Principal component analysis, Linear discriminative analysis, Multiple discriminant analysis, Independent component analysis.
	Reinforcement Learning [2Hrs] Markov Decision Process, Planning, Estimation, Control and Applications.
	Recommendation [3 Hrs] Systems Introduction, Types of recommender systems, Content-based, Collaborative filtering: Matrix factorization based approaches, Knowledgebase, and Hybrid techniques, Times series forecasting, other real time examples.
	Hands-on to the majority of the topics using R.(Practical Implementation)[08 Hrs] **Projects**

	> Diabetic's prediction using KNN. > Color compression using K-means clustering > Times series forecasting > Student performance prediction using decision tree
Total Number of Session	40 HRS
Teaching Faculties	1)Prof.Rashmi Varade(course coordinator) 2)Dr.Yugandhara More 3)Dr.Janhavi Raut 4)Prof.Jagruti Raut(HOD dept of CS Viva college,Virar) 5)Prof.Poornima Rane(N.b.Mehta College,Bordi)

Rashmi Varade
Incharge CS Dept

Prof. Mahesh Deshmukh
(IQAC Co-ordinator)

Dr. Kiran J. Save

List of the students enrolled in the course

Sr.no.	Name of a student	Class:-	Contact no:-	Email Id of student
1	Akshay Singh	TYBSc	8830873847	akshayhajari2108@gmail.com
2	Aniket Jaiprakash Raut	Sycs	8408022086	aniketraut6631@gmail.com
3	Ansh Patil	FY.bsc.CS	9607447406	Anshpatil04@gmail.com
4	Asif munawwar khan	Sy bsc Cs	7709532534	ak4895925@gmail.com
5	Bhargav Mali	Ty Bsc cs	9834438669	Bhargav.mali1301@gmail.com
6	Bind Abhishek Somaru	S Y Bsc CS	9503283349	bindabhishek007@gmail.com
7	Deep Bharat kadu	Sycs	9172874727	kadudeep0@gmail.com
8	Durgesh yashwant sawant	T.y.b.sc.C.S	9325491447	durgesh2001sawant@gmail.com
9	Gaurav Desale	Ty-Bsc	8888672520	gaurav_desale@yahoo.com
10	Kavita Santya Bendga	S.Y.B.Sc CS	7498197886	kvbendga1012@gmail.com
11	KOMAL RAVINDRA CHAUDHARY	Sybsc CS	8766840003	kchaudhari392@gmail.com
12	Mitra Tiwari	TY.Bsc(CS)	9168257414	mitratiwari99005@gmail.com
13	Nitisha Pankaj Rane	TY CS	7262836747	nitisharane8@gmail.com
14	Payal Dilip Rahane	Bsc CS (TY)	9970864474	payalrahane965@gmail.com
15	Payal Gayaprasad Gond	Bcs cs	9175798631	Payalgond17@gmail.com
16	Pooja ramprasad yadav	Tybsc CS	9727690806	py788370@gmail.com
17	Pratap Mohan Bhool	Sy Bsc CS	7769013954	pratapbhool377@gmail.com
18	Priya Hiraman Rajak	Tybsc cs	8208545002	Priayrajak912@gmail.com
19	Rahul Dhokane	TYCS	8208672101	rahuldhokane002@gmail.com
20	Rai sonam tarkeshwar	Sycs	9130295366	sonamr2908@gmail.com
21	Ritvik Bhupesh Mali	Sybsc cs	9325942950	ritvikmali2@gmail.com
22	Rohini narsingrao ambegave	Ty bsc cs	8793618348	rohininarsingrao@gmail.com
23	Sanyukta Deepak Patil	Sy bsc cs	8010598796	sanyuktapatil2002@gmail.com
24	Satnam Kour	Ty cs	8007730666	satnamkour7@gmail.com
25	Shaw Rohitkumar Nagendra	SY BSC CS	7262900629	shawrohit7262900629@gmail.com
26	Shivangi	Sybsc cs	8655295128	Shi1vangi20@gmail.com
27	Shree Ram Sanat Mishra	Sycs	7972523565	mishraram7972@gmail.com
28	Sudip Tapas Hajra	BSC CS	9834298051	sudiphajra222@gmail.com
29	Suraj kailas gaikwad	Sybsc CS	9156989995	surajgaikwad14337@gmail.com
30	SUYASH RAMESH DAPALE	FYBSC CS	8511964216	daplebunty143@gmail.com
31	Swapnali bharat patil	Tycs	7218057167	Patilswapnali418@gmail.com
32	Swati Harishchandra Gupta	SY CS	9767596832	swati.shyama@rediffmail.com
33	Tanishk Mestry	1st year bsc	8087594933	tanishkmestry4183@gmail.com
34	Vaibhav Shinde	Bsc cs 3 sem	8208019055	vaibhavss808@gmail.com
35	Vaishnavi thorat	Sycs	7796141433	Vaishuthorat2003@gmail.com
36	Vinit Tiwari	Tybsc (cs)	8459696073	tvinit3105@gmail.com
37	Dipti Tripathi	SyCS		tripathidipti496@gmail.com

Certification Course on Machine Learning for Data Science using R

No. of questions: 50   **Total Marks: 50**

1. Clustering techniques are _____ in the sense that the data scientist does not determine, in advance, the labels to apply to the clusters. *
   - Unsupervised
   - Supervised
   - Reinforcement
   - Neural network

2. Euclidean distance measure is *
   - A stage of the KDD process in which new data is added to the existing selection.
   - The process of finding a solution for a problem simply by enumerating all possible solutions according to some pre-defined order and then testing them
   - The distance between two points as calculated using the Pythagoras theorem
   - none of above

3. Classification accuracy is *
   - A subdivision of a set of examples into a number of classes
   - Measure of the accuracy, of the classification of a concept that is given by a certain theory
   - The task of assigning a classification to a set of examples
   - None of these

4. Which of the following algorithm is most sensitive to outliers? *
   - K-means clustering algorithm
   - K-medians clustering algorithm
   - K-modes clustering algorithm
   - K-medoids clustering algorithm

5. Hidden Markov Model is used in- _____ *
   - Supervised learning
   - Unsupervised learning
   - Reinforcement learning
   - All of the above

6. Which of the following techniques would perform better for reducing dimensions of a data set? *
- Removing columns which have too many missing values
- Removing columns which have high variance in data
- Removing columns with dissimilar data trends
- None of these

7. Which of the following function is used for k-means clustering? *
- k-means
- k-median
- heatmap
- none of the mentioned

8. The square root of the variance is called the _____ deviation. *
- mean
- standard
- continuous
- empirical

9. Which of the following is a lazy learning algorithm?
- SVM
- KNN
- Decision tree
- RF

10. What does k stand for in the KNN algorithm? *
- Number of neighbors
- Number of output classes
- Number of input features
- None

11. With the help of a confusion matrix, we can compute-
- Recall
- Precision
- Accuracy
- All of the above

12. Machine learning algorithms build a model based on sample data, known as .... *
- Training Data

- Transfer Data
- Testing Data
- None of the above

13. Machine Learning is a field of AI consisting of learning algorithms that .... *
- At executing some task
- Over time with experience
- Improve their performance
- All of the above

14. The discrete variables and continuous variables are two types of _____. *
- Open end classification
- Time series classification
- Qualitative classification
- Quantitative classification

15. What is the minimum no. of variables/ features required to perform clustering? *
- 0
- 1
- 2
- 3

16. Reinforcement learning is- _____
- Unsupervised learning
- Supervised learning
- Award based learning
- None

17. Which of the following is required by K-means clustering? *
- defined distance metric
- number of clusters
- initial guess as to cluster centroids
- all of the mentioned

18. False negatives are- *
- Predicted negatives that are actually positives

- Predicted positives that are actually negatives
- Predicted negatives that are actually negatives
- Predicted positives that are actually positives

19. Decision trees use _____ , in that they always choose the option that seems the best available at that moment. *
- Shortest path algorithm
- Backtracking
- divide and conquer
- Greedy Algorithms

20. .... is a widely used and effective machine learning algorithm based on the idea of bagging. *
- Regression
- Classification
- Decision Tree
- Random Forest

21. Bayesian classifiers is0 *
- A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory
- Any mechanism employed by a learning system to constrain the search space of a hypothesis
- An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation
- None of the above

22. What are the three types of Machine Learning? *
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- All of the above

23. Which of the following is the valid component of the predictor? *
- data
- question
- algorithm
- all of the mentioned

24. Which of the following is correct order of working? *
- questions->input data ->algorithms
- questions->evaluation ->algorithms
- evaluation->input data ->algorithms
- all of the mentioned

25. Which algorithm is used for solving temporal probabilistic reasoning? *
- Hill-climbing search
- Hidden markov model
- Depth-first search
- Breadth-first search

Sonopant Dandekar Shikshan Mandali College, Palghar
Department of Computer Science

Sr. No.	Topics Covered	Date	Session Hours
1	R- Installation	8/3/2021	1
2	Introduction	8/4/2021	1
3	Machine Learning Types	8/5/2021	1
	Total		3

Dr. Yugandhara More

Sonopant Dandekar Shikshan Mandali College, Palghar
Department of Management Studies

Sr. No.	Topics Covered	Date	Session Hours
1	R programming Basics	10/20/2021	2
2	Assigning values to variables, Basic Data Types	10/21/2021	2
3	Classification	10/22/2022	2
4	Decision Tree, KNN	10/23/2022	2
5	Clustering	10/25/2022	2
6	Probability Distribution	10/26/2022	2
7	Application of clustering	10/26/2022	2
8	Diabetic's prediction using KNN	10/27/2022	2
9	Color compression using K-means clustering	10/28/2022	2
10	Times series forecasting	11/24/2022	2
11	Student performance prediction using decision tree	11/29/2022	2
	Total		22

Dr. Janhavi Raut

Sonopant Dandekar Shikshan Mandali College, Palghar
Department of Management Studies

Sr. No.	Topics Covered	Date	Session Hours
1	Dimensionality Reduction	1/19/2022	2
2	Time series forecasting	1/19/2022	1
3	Reinforcement Learning	1/20/2022	2
4	Recommendation system	1/28/2022	2
5	Times series forecasting	1/30/2022	2
6	Regression Analysis	1/31/2022	2
7	Polynomial Regression	2/7/2022	2
8	Probability and statistics for data science	2/12/2022	2
	Total		**15**

Prof. Pournima Jawale