DATA MINING AND BIG DATA ANALYTICS

UNIT 1

Concepts – Data mining process – Data preprocessing - Exploratory Data Analysis - Mining Frequent Patterns- Associations and correlations.

Classification and Prediction

Classification by decision tree – Rule-based classification – Support vector machines – Associative classification – k- Nearest-Neighbor classifiers – Prediction: Accuracy and error measures- Evaluating accuracy of a classifier and predictor- Ensemble methods – Examples.

UNIT 2

Clustering

Cluster analysis – Partitioning methods – Hierarchical methods – Density based method – Grid based method – Outlier analysis – Examples.

ANN and applications

Introduction to artificial intelligence- Concepts of neural networks -Developing NN-based systems- Choice of network structure- Applying and training neural networks- other NN paradigms- Genetics Algorithm – Fuzzy Logic – Examples.

UNIT 3

Text (unstructured) Mining

Information retrieval- Using Text for Prediction- Textual Information to Numerical Vectors dimensionality reduction for text- Introduction to NLP - Web content Mining, Web Structure Mining, Web usage Mining. Spatial Mining – Temporal Mining- Examples

UNIT 4

Introduction to Big Data Analytics & Data Science

What is big data-Why big data matters- Big data and business case-Big data sources- Web scraping, crawling, crowdsourcing, crowdsensing, Big data technologies and platforms, NOsql and map-reduce paradigm, Machine data Vs Human generated data -drivers of big data-Analytic data sets- Big data dimensions-Volume, variety and velocity- Industry examples of big data- impact of cloud computing- Evolving concept of Data Science-learning

84

from knowing-agility-multidisciplinary focus- scale & convergence leading to Data Science-Data Scientist-How is a data scientist different from a Statistician

Theories and Methods

The evolution of big data analytics- Look (Search, Indexing and Memory) - Listen (Streams, Information and Language, Analyzing Sentiment and Intent) – Learn (Classification, Clustering, and Mining, Information Extraction) - Connect (Reasoning: Logic and its Limits, Dealing with Uncertainty)- Predict: Forecasting, Neural Models, Deep Learning- Collaborative filtering-Large graph analysis-Text mining- Volume Trending- Influencer Identification- In-Memory Analytics.

UNIT 5

Applications of Big data

Listening to social media sources - Monitoring social trends - Basics of opinion mining and sentiment analysis- Exemplar social media mining projects - Big data proxies of human mobility - Basic measures of human mobility. Data-driven human mobility models - Mobility data mining with GPS tracking data - Analysis of traffic and city dynamics with vehicular telematics data - Analysis of personal vs. collective mobility

Text /References:

1. Michael Minelli, Michele Chambers, Ambiga Dhiraj ,"*Big Data, Big Analytics*", John Willey , 2013
2. Chris Eaton, Dirk Deroos, Tom Deutsch, George Lapis, Paul Zikopoulos, "*Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*", Tata McGraw Hill Education, 2012
3. Han, J and Kamber M., "*Data mining-concepts and applications*", 3rd Edition, Elsevier publications, 2011
4. Margaret H. Dunham, "*Data Mining : Introductory and Advanced Topics*", Pearson Education, First Indian Reprint,2003
5. Weiss, S.M., Indurkhya, N., Zhang, T., Damerau, F., "*Text Mining, Predictive Methods for Analyzing Unstructured Information*", Springer publications, 2005
6. Anthony Scime, "*Web mining: applications and techniques*", Idea Group publications, 2007
7. Paolo Giudici, "*Applied Data mining: Statistical methods for business and industry*", John wiley & sons(Asia) pte Ltd., 2005

Senate
RU