Name of the Course:	**TEXT ANALYTICS**
Name of Instructor:	Dr Santosh Kumar Vishwakarma
Language of Instruction:	English
Number of Contact Hours:	36
Credit Awarded:	03

Objective of course:

Perform all common data preparations

Build strong analytical predictive models

Evaluate model quality with respect to different performance criteria

Deploy analytical predictive models

Identify techniques for processing unstructured data

Transform textual data into a structured format and perform necessary pre-processing

Apply different statistical text-processing methods

Syllabus:

Fundamentals of CRISP-DM

EDA: Exploratory Data Analysis

Data Preparation

Predictive Modelling Algorithms

Model Construction and Evaluation

Machine Learning - Variance, Overfitting & Underfitting

Text Processing - Visualization & Pre-processing of Textual Data

TF-IDF & its variation methods

Advanced Modelling Methods

Web Mining

Organization of the Course:

Total contact hrs 36		
1st week:	10 hrs (classes)	2 hrs (self-study/project)
2nd week:	10 hrs (classes)	2 hrs (Mid term exam/assessment/discussion)
3rd week:	10 hrs (classes)	2 hrs (End term exam)

Mode of lectures: online lecture/online videos/case study/ discussion/ workshop/ hands-on

Course Plan:

Lecture no.	Topic	Lecture mode	Instructor
L: 1-6	Overview - Analytics Taxonomy, CRISP-DM Getting Started with RapidMiner Studio - User Interface, Creating and Managing RapidMiner Repositories, Operators and Processes, Storing Data, Processes, and Result Sets EDA: Exploratory Data Analysis, Loading Data Quick Summary Statistics, Visualizing Data & Basic Charting Data Preparation - Basic Data ETL, Data Types & Transformations of Value Types, Handling Missing Values, Handling Attribute Roles, Filtering Examples and Attributes, Normalization and Standardization	Online with Hands-on	Dr Santosh K Vishwakarma
L: 7-12	Predictive Modelling Algorithms - k-Nearest Neighbor, Naïve Bayes, Linear Regression, Decision Trees & Rules Model Construction and Evaluation - Machine Learning Theory: Bias, Variance, Overfitting & Underfitting, Splitting Data, Split and Cross Validation, Evaluation Methods & Performance Criteria	Online with Hands-on	Dr Santosh K Vishwakarma
L: 13-18	Loading of Texts - Loading from Flat Files, Data Sets, Web Sources (e.g. URL crawling, Twitter) Text Processing – Visualization - Visualizing Documents and Tokens, Multi-Dimensional Visualizations, Handling Unstructured Data, Pre-processing of Textual Data, Tokenizing, Stemming, Filtering of Tokens Case Sensitivity, Term Frequencies, Document Frequencies, TF-IDF	Online with Hands-on	Dr Santosh K Vishwakarma
L: 19-24	Advanced Modelling - Support Vector Machines, Naive Bayes	Online with Hands-on	Dr Santosh K Vishwakarma
L: 25-30	Text Clustering Methods, Web Mining - Crawling the Web, Extracting Information from Web Sites, Transforming Web Sites to Documents, Retrieving Structured Web Data	Online with Hands-on	Dr Santosh K Vishwakarma

Brief Profile of the Instructor

Dr. Santosh K. Vishwakarma has 10+ years of experience in the field of Information Retrieval & Data Analytics. He is working as Associate Professor in School of Computing & Information Technology at Manipal University Jaipur. His teaching specialization includes database management system, Information Retrieval, Machine Learning, Data Analytics. His research interest area includes term weighting models in information retrieval, static index pruning algorithms, predictive analysis, trend detection, regression, classification and clustering algorithms in data mining and text mining.

He has delivered lectures in various National & International forums including the RapidMiner World Conference 2015 in Boston, USA.