CERTIFICATE COURSE WITH DATA SCIENCE WITH

3.Natural Language Processing and Big Data

UNIT-I

Introduction to NLP.

What is NLP, Various levels of NLP: Morphological,Lexical Analysis, Syntactic analysis, Semantic analysis,Discourse level, Pragmatic),Applications of NLP

Introduction to Text Processing: Working with,Text Files, HTML files, XML files,JSON files and PDF files,Working with Regular Expressions

UNIT-II

Text Processing using NLTK, Blob, Spacy

Text Processing: Tokenization,Stemming, Lemmatization, Removal of Stop Words, POS tagging and Named Entity recognition, Text Preprocessing, Phrase Matching

Text Feature Extraction using SciKit-Learn: Vector Space Model representation, Term Frequency, Document Frequency, TF_IDF frequency, Count Vectorizer, TF-IDF Transformer, TF-IDF Vectorizer, Text Similarity

UNIT-III

Application Development using Text using ML

Text Classification,

Text Clustering and

Text Summarization

Case Studies and Application development

Topic Modelling using NLP: Introduction to Topic Modelling, Latent Dirichlet Allocation with Python - Part Two, Case studies and Applications

Sentiment Analysis: Introduction to Sentiment Analysis

Creating NLP Pipeline for Text Mining (Social Media data/Web data), Word2Vec and Doc2Vec, Transformers, Recommendation Systems - Collaborative filtering, Overview of Language Modelling

UNIT-IV

Introduction to Big Data,Evolution of Bigdata, Types of Digital data, Characteristics& Challenges of data, Overview of Predictive Analytics, NoSQL databases

UNIT-V

Key Technologies and Drivers for Big Data

Knowledge Discovery Tools, Stream Analytics, In-memory Data Fabric, Distributed Storage and Computing, Data Integration and Visualization, Data Pre-processing

UNIT-VI

Hadoop Eco System

Hadooop for Bigdata, Overview of Apache Hadoop software, Installation of Hadoop, Architecture of Hadoop, Understanding Hadoop eco-system-HDFS, Map Reduce, Working with Hadoopeco system components- Hive, Pig, Data Ingestion with Flume &Sqoop,HBase

UNIT-VII

Bigdata&In-memory computing

Understanding In-memory computing, Resilient Distributed Databases(RDDs), Introduction to Big Data Analytics with Spark, Understading Spark eco-system components, Overview of client mode & cluster mode computing, Working with basic Spark scripts, Data Analytics using Spark eco-system

Case Studies & Applications of ML in Spark

UNIT-VIII

Real-time Streaming platforms for Big Data

Overview of Apache Kafka & Storm