Master of Engineering - ME (Big Data and Data Analytics)

The recent years have witnessed a tremendous technological growth in managing, organizing and harnessing the power of large scale data. Big Data and Data analytics are playing an important role in defining the vision in every sphere of life. There is no human institution not influenced by big data and analytics. Business, government, healthcare, education and the society as a whole, derives insights from the historical data. Analytics helps in predicting potential opportunities as well as predicting the possible future.

Big Data is about providing efficient technological “solution stacks” to organize and access large scale data. Data Analytics is about combining principles and techniques from mathematics, computer science and machine learning to predict possibilities as well as to prescribe actions. The Big Data and Data Analytics program offers a well-balanced content organized across one year course work. The students get to learn first principles of Data Analytics – statistics, probability, machine learning, and text analytics along with the technological foundations of Big Data frameworks. Since every course is supplemented by a laboratory component, students get to learn and acquire skills that are crucial in the field. This program prepares students to confidently cope with the ever growing demands of this exciting field.

Program Highlights:

1. **Practical**: All the subjects are associated with lab to provide the practical knowledge.

2. **Mini Project**: Two Mini Projects during the program. In this students are expected to select a problem in the area of their interest and the area of their specialization that would require an implementation in hardware / software or both in a semester.

3. **Seminar**: In this students have to make a literature survey and select a latest topic in the area of their interest and the area of their specialization and make a presentation in the semester.

4. **Project**: Internship project at the Industry/Institution for 10 to 12 months.

<table>
<thead>
<tr>
<th>Semester - I</th>
<th>Semester - II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithms and Data Structures for Big Data</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>Large Scale Distributed Computing Systems</td>
<td>Architecture of Big Data Systems</td>
</tr>
<tr>
<td>Probability and Statistical Inferences</td>
<td>Multiple Linear Regression and Logistic Regression</td>
</tr>
<tr>
<td>Modern Database Management Systems</td>
<td>Healthcare Informatics</td>
</tr>
<tr>
<td>Elective – 1</td>
<td>Elective – 2</td>
</tr>
</tbody>
</table>

**Elective subject details**

<table>
<thead>
<tr>
<th>Elective - 1</th>
<th>Elective - 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DevOps for Big Data Systems</td>
<td>Text Retrieval and Search Engines</td>
</tr>
<tr>
<td>Mobile Web Application Development</td>
<td>Applied Multivariate Analysis</td>
</tr>
</tbody>
</table>
Semester III & IV

Internship / Project work for 10 – 12 months in Industry / Research Organization / University.

**Brief Syllabus**

**Semester - I**

1. **Algorithms and Data Structures for Big Data**
   Algorithm specification and analysis techniques, Elementary data structures, Sorting and Searching Techniques, Hashing and Dictionaries, Binary search trees, Graphs, String and text processing techniques, Data stream algorithms.

2. **Large Scale Distributed Computing Systems**
   Introduction to fundamental concepts, Architectures, Processes, Communication, Synchronization - election algorithms, Consistency, replication and recovery, Security, Distributed Web-based systems.

3. **Probability and Statistical Inferences**
   Applied probability, Random variables and distributions, Point and interval estimation, Hypothesis testing.

4. **Modern Database Management Systems**
   Introduction to the growth of traditional and modern database management systems, Foundations of relational data management, SQL - syntax and semantics, Designing relational data - normal forms, Transaction processing, Two-Phase commit protocol, Semi-structured data management, NoSQL – origins, growth and applications, Distribution models for scalability, MapReduce, UnQL: a query language and algebra for semi-structured data based on structural recursion, NewSQL

5. **Elective – I**

**Semester – II**

1. **Machine Learning**
   Introduction, Inductive Classification, Predictive analytics – Supervised learning, Ensemble methods (bagging and boosting), Computational learning theory, Bayesian learning, Instance-based learning, Support Vector Machine (SMV), Descriptive analytics – unsupervised learning, Clustering.

2. **Architecture of Big Data Systems**
   Classifying big data characteristics, Big Data processing - the Lambda architecture, Data storage on the batch layer, Computing on the batch layer, Serving layer, Speed layer, Alternatives to MapReduce
3. Multiple Linear Regression and Logistic Regression

Linear regression, Logistic regression.

4. Healthcare Informatics

Theoretical foundations of health informatics, Evidence-based practice and informatics, Electronic health records and managing patient care, Telehealth and applications for delivering care at a distance, Imaging Technologies and their Applications in Biomedicine, Clinical decision support systems in healthcare, Public health informatics, Participatory healthcare informatics, Privacy, confidentiality, security and data integrity, Patient Safety and Quality Initiatives in Informatics, Improving user experience for health information technology products, Standards in Healthcare industry and practice, Distance Education: Applications, Techniques and Issues, Information Systems and Technical Tools in Healthcare Education

Elective – II

Elective - I

DevOps for Big Data Systems

Agile Infrastructure management and DevOps, Scripting with Python and Linux Shell, SSH and centralized authentication/authorization, Version control systems and Configuration Management Systems, The Hadoop environment, HTML5 and JavaScript for DevOps, Web Applications, testing and performance analysis,

Mobile Web Application Development

Challenges of mobile Web application development, Setting up a personal Web site, HTML5 and CSS for mobile devices, Programming with JavaScript and DOM APIs, Architecture of Android applications, Programming for technologies available on smart phones, Developing offline facilities in mobile web applications, Designing and developing secure mobile web applications

Elective – II

1. Text Retrieval and Search Engines

Text and Web search basics, Architecture of a Search Engine, Processing Text, Queries and Interfaces, Boolean Retrieval, Scoring, term weighting and the Vector Space Model, Probabilistic Models, Text classification & Naive Bayes, Social Search

2. Applied Multivariate Analysis

Multivariate data and multivariate normal distribution, Multivariate test on mean vectors and discriminant analysis Principal Component Analysis and factor analysis, Cluster Analysis,