# ATRIA INSTITUTE OF TECHNOLOGY

**Proposal by**

## DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

**To Have Center of Excellence**

**On**

# DATA ANALYTICS

DATA ANALYTICS CENTER OF EXCELLENCE

# Table of Contents

# 1. VISION AND MISSION

## VISION OF ISE DEPARTMENT

To be the centre of excellence in Information Science and Engineering for enabling students to fulfil their aspirations by imparting **quality education** to **explore emerging trends, techniques and technologies** and ***encouraging inquisitiveness, curiosity*** **and the thirst to learn new things** in students to meet the **needs of the society and industry.**

## VISION OF THIS CoE

To be the centre of excellence in ***Data Analytics*** for ***encouraging inquisitiveness & curiosity*** in students and **to prepare students to be experts** in **solving business problems** to help **business executives in decision making,** by getting well-versed with well-known **data analytics tools and techniques.**

## MISSION OF THIS CoE

- **To prepare students to be Data Analytics experts:** To prepare students for entry-level data analyst jobs.
- **In solving business problem:** Mapping and tracing the data to solve business problems.
- **To help business executives in decision making:** Designing and creating the data-reports to help business executives in decision making.
- **Data analytics tools and techniques:** Training and mentoring students on popular analytical tools which is an add-on to their CVs to make high calibre and work-ready data analytics graduates.

# 2. OVERVIEW

## 2.1 The Big Data

- Big data can be applied to real-time fraud detection, complex competitive analysis, call centre optimization, consumer sentiment analysis, intelligent traffic management, and to manage smart power grids, to name only a few applications.

- Big data is characterized by three primary factors: volume (too much data to handle easily); velocity (the speed of data flowing in and out makes it difficult to analyze); and variety (the range and type of data sources are too great to assimilate).

- With the right analytics, big data can deliver richer insight since it draws from multiple sources and transactions to uncover hidden patterns and relationships

- There are four types of big data BI that really aid business:

    1. **Prescriptive** – This type of analysis reveals what actions should be taken. This is the most valuable kind of analysis and usually results in rules and recommendations for next steps.

    2. **Predictive** – An analysis of likely scenarios of what might happen. The deliverables are usually a predictive forecast.

    3. **Diagnostic** – A look at past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.

    4. **Descriptive** – What is happening now based on incoming data. To mine the analytics, you typically use a real-time dashboard and/or email reports.

## 2.2 Big Data Analytics in Action:

Big data analytics can deliver big value to business, adding context to data that tells a more complete story. By reducing complex data sets to actionable intelligence, we can make more accurate business decisions.

1. **Prescriptive analytics:** is where big data analytics in general sheds light on a subject, prescriptive analytics gives you a laser-like focus to answer specific questions.

2. **Predictive analytics:** use big data to identify past patterns to predict the future.

3. **Diagnostic analytics**: are used for discovery or to determine why something happened.

4. **Descriptive analytics:** In other words, data mining are at the bottom of the big data value chain, but they can be valuable for uncovering patterns that offer insight

CoE-Data Analytics will focus on open source software and platforms as much as possible to stay nimble, foster efficiency and drive innovation.
It has to provide access to technologies in the following areas:

- Data collection Data storage
- Data curation and management
- Data processing
- Data exploration and visualization
- Unstructured data processing
- Data analytics and machine learning

## 2.3 Technologies Required to learn Big Data Analytics:

- **Hadoop Platform:** (Big Data Processing and analytics)

  → To Processing Big Data on Hadoop, understand how to use Hadoop to handle and arrange Big Data.

  → To teach how to perform analytical operations to gain insights from data processed through Hadoop.

- **Why is Big Data Analytics on Hadoop important?**

  → Businesses are now aware of the large volumes of data that they generate in their day to day transactions. They have also realized that this Big Data can provide very valuable insights once analysed.

  → The massive volume of Big Data and its unstructured format make it difficult to analyse Big Data. Hadoop brings the ability to cheaply process large amounts of data, regardless of structure.

  → Knowledge about Big Data Analytics on Hadoop will also prove to be a huge resume builder for Students who are aiming to work in the IT Industry

- **Different Hadoop Platforms:**

  → Apache, Horton work, Cloudera.

  → Tools Taught: Map-Reduce Programming, hive, Pig, NoSQL, Flume, Hue, HBase, Cassandra, Yarn.

- **Integration of Hadoop:**

  → With R: R-Hadoop, Python and Mahout

  → Level 1: Hadoop Platform: processing large volume of data using cluster techniques and convert unstructured, semi structure data into structure form.

  → Level2: Java, Python or R:

  → Level 3: ML

→ Level4: Visual Analysis tools: Tableau, Python with Pandas, tense flow, R with libraries

- **Paid Tools:**

  → Tableau, Microsoft azure, IBM Watson.

- **Associated Technologies:**

  → Cloud and IoT (Sensor data), Web App, Android App

# 3. EXECUTION PLAN

| PHASE | DESCRIPTION | DURATION | TARGET STUDENTS |
|---|---|---|---|
| 1 | Start with Industry Experts (Importance of Data, use cases and Future Trends) | February 2019, 1st week | 3rd Semester and 5th Semester |
| 2 | Deliver session on Hadoop Platform with all possible tools. | February 2019 - March 2019 | 3rd Semester and 5th Semester |
| 3 | Deliver sessions on Python Language Knowledge transfer to students through Project exhibition | April 2019 – May 2019 | 8th Semester |
| 4 | Deliver sessions on Machine Learning | August 2019 – October 2019 | 3rd Semester and 5th Semester |
| 5 | Assign projects to CoE-registered students and take up consultancy projects | | 3rd Semester and 5th Semester |

# 4. ASSOCIATED FACULY

| | | |
|---|---|---|
|  | Name of the Faculty | **Dr. Neha Mangla** |
| | Designation | **Associate Professor** |
| | Department | **Information Science and Engineering** |

| | | |
|---|---|---|
|  | Name of the Faculty | **Abhilash** |
| | Designation | **Assistant Professor** |
| | Department | **Information Science and Engineering** |

| | | |
|---|---|---|
|  | Name of the Faculty | **Ayesha Tarannum** |
| | Designation | **Assistant Professor** |
| | Department | **Information Science and Engineering** |

# 5. ASSOCIATED STUDENTS

| SL.NO | STUDENT NAME | SEMESTER |
|---|---|---|
| 1 | Monica S Kumar | 7th |
| 2 | Namitha Hegde | 7th |
| 3 | Sushmitha | 7th |
| 4 | Priya B | 7th |
| 5 | Swathi K Bhat | 7th |
| 6 | Sachin | 7th |
| 7 | Kulsum | 7th |
| 8 | Mahima Krishna | 7th |
| 9 | Swetha Ramesh | 7th |
| 10 | Fariya Banu | 7th |
| 11 | Prakash Choudary | 7th |
| 12 | Prathiksha Pole | 7th |
| 13 | Niveditha | 7th |
| 14 | Simran | 7th |
| 15 | Abhinandan | 7th |
| 16 | Mayank | 5th |
| 17 | Neha Prakash | 5th |
| 18 | Ganesh A | Recently Passed Out |
| 19 | Narayana | Recently Passed Out |

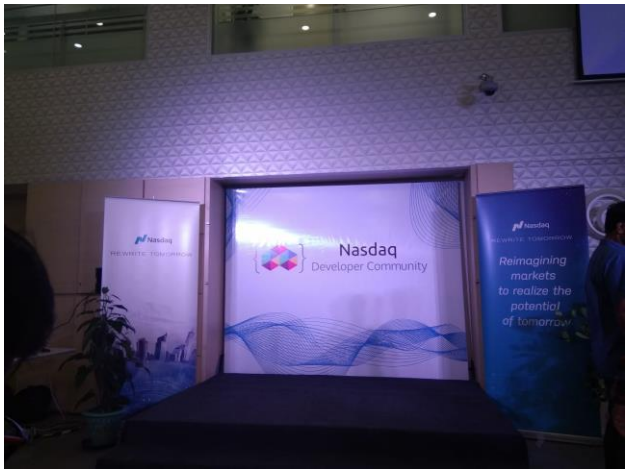
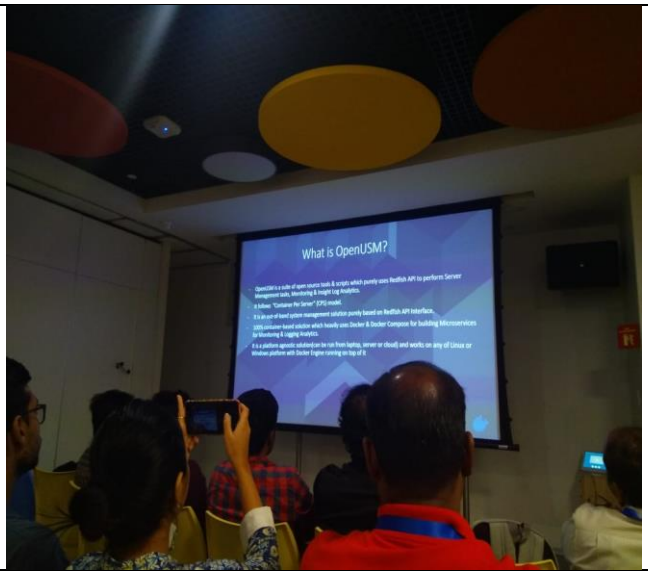
# 6. SYLLABUS INCLUDED

| PART - 1 | |
|---|---|
| **Syllabus** | **Topics** |
| Knowledge on Hadoop which includes | Hadoop: Map-Reduce Framework, essential Hadoop Tools. |
| Data Mining Concepts | Association analysis, classification, Clustering |
| Knowledge on python language | String list, dictionary, numpy, scikit learn, pandas. |
| Projects using python based on mentioned data mining concepts | |

| PART - 2 | |
|---|---|
| **Syllabus** | **Topics** |
| More advanced concepts in machine learning | Decision tree learning, ANN-based learning, Bayesian learning |
| Probability and distributions | Algorithms |
| Projects based on mentioned concepts | |

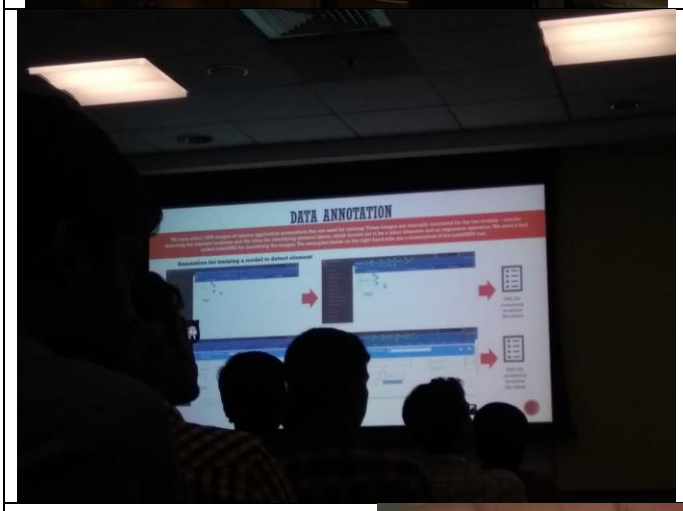
| PART – 3 |
|---|
| **Project based Learning** |
| Implementing real time projects |
| Consultancy Projects |
| Workshops for other departments |

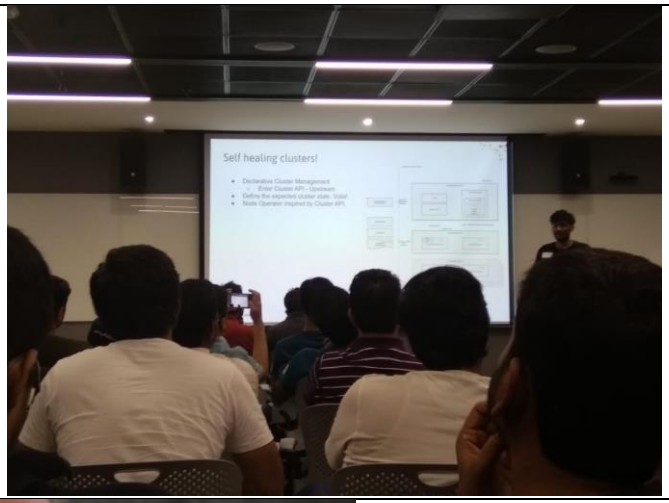
# 7. COE STUDENTS ATTENDING TECHNICAL MEET-UPS

<table>
<tr>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td> 
</td>
<td> 
</td>
</tr>
</table>


Redfish Operational Model


IBM
Center for Learning and Development


DATA ANNOTATION


Self healing clusters!


Google Cloud ML Study Jam
India - 2019
SACHIN
#DevCommunity
#GDG
9th Feb 2019



# DEPARTMENT OF INFROMATION SCEINCE AND ENGINEERING

## CENTRE OF EXCELLENCE IN DATA ANALYTICS

# VISION

- To be the centre of excellence in ***Data Analytics*** for ***encouraging inquisitiveness & curiosity*** in students and **to prepare students to be experts** in **solving business problems** to help **business executives in decision making,** by getting well-versed with well-known **data analytics tools and techniques.**

# MISSION

- **To prepare students to be Data Analytics experts:** To prepare students for entry-level data analyst jobs.
- **In solving business problem:** Mapping and tracing the data to solve business problems.
- **To help business executives in decision making:** Designing and creating the data-reports to help business executives in decision making.
- **Data analytics tools and techniques:** Training and mentoring students on popular analytical tools which is an add-on to their CVs to make high calibre and work-ready data analytics graduates.

# MOTIVATION

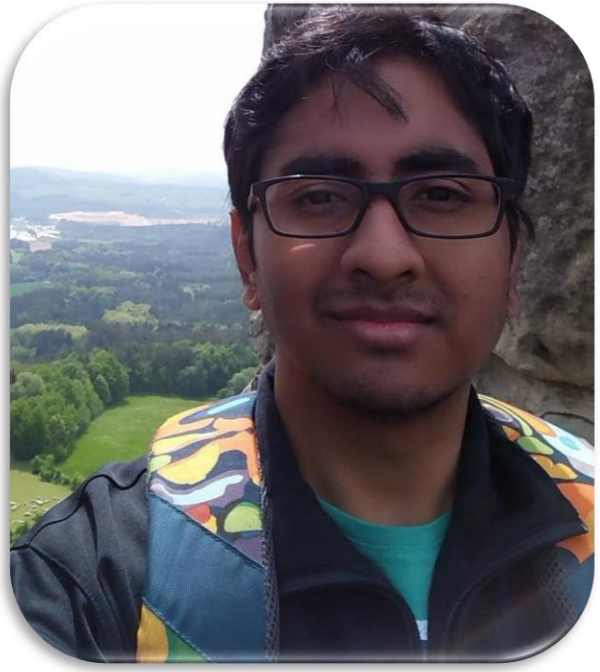
- To be a well known Department for Data Analytics
- Atria 2.0 Framework
- Department re-framed vision and mission

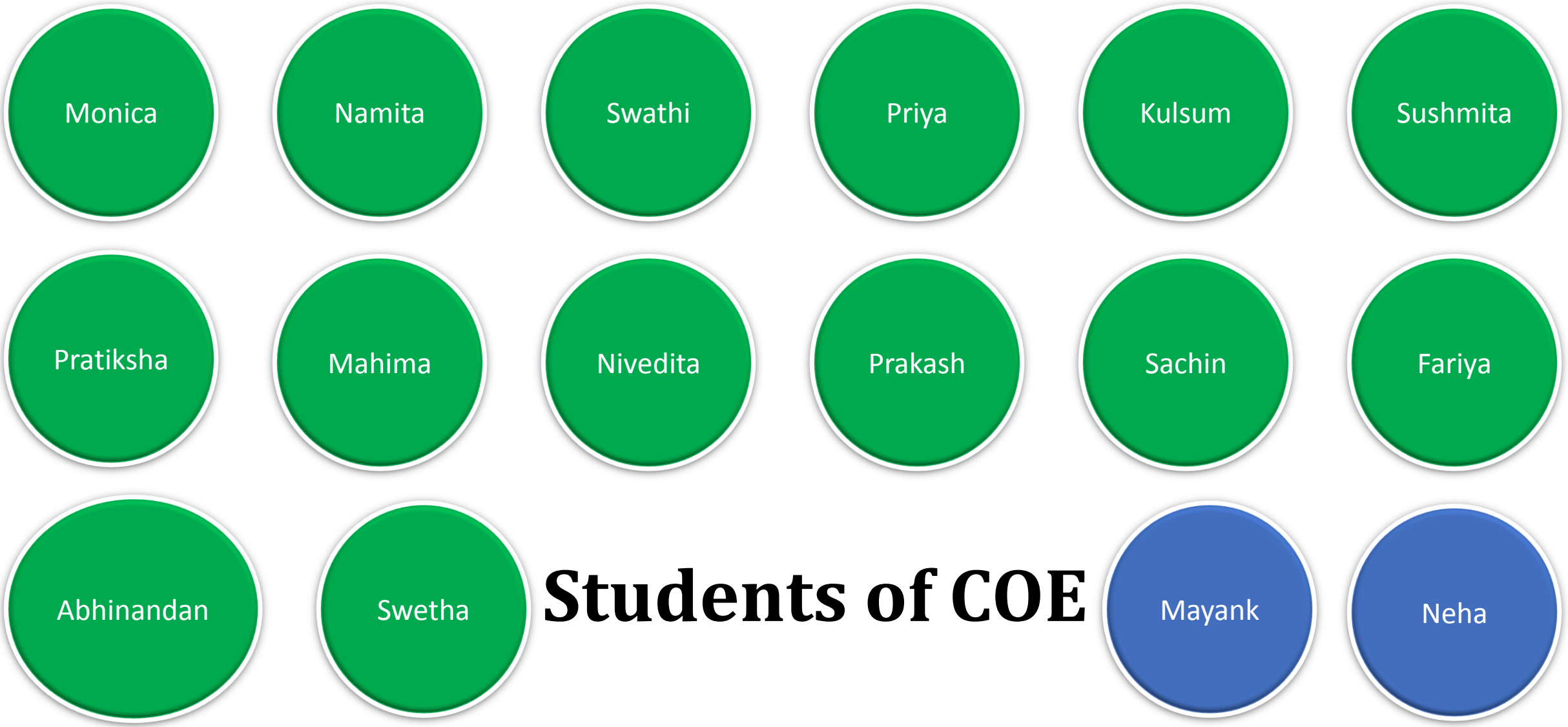
# Why we chose CoE in Data Analytics?

- The domain is very vast and trending
- Very much beneficial in terms of jobs with a decent package
- ***Support System:*** Interest and knowledge of teachers on required subjects that are necessary to understand data analytics.
- Faculty Coordinators:
  - ❖ Dr. Neha Mangla: Python Programming, Machine Learning
  - ❖ Mrs. Ayesha Taranum: Data Mining and Data Warehousing
  - ❖ Mr. Abhilash: Big Data and its Significance, Machine Learning

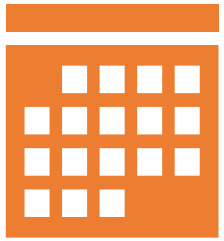
# Why we chose CoE in Data Analytics?

- Alumni Coordinator: **Dr. Prajit Dhar,**
  - Alumni of ATRIA (2012 batch pass out)
  - Education:
    - Master of Science in Computational Linguistics, University of Stuttgart, Germany (2015-2017)
    - Studied Language, Logic and computation, University of Amsterdam (2016-2017)
  - Work:
    - Former Research assistant, at Fraunhofer IPA, Stuttgart, Germany
    - UrbanPro.com , Corporate Trainer (2014-2015)
    - Amazon.com, Transaction Risk Investigator, (2012-2014)




Monica
Namita
Swathi
Priya
Kulsum
Sushmita
Pratiksha
Mahima
Nivedita
Prakash
Sachin
Fariya
Abhinandan
Swetha
Students of COE
Mayank
Neha

# When Did We Start?

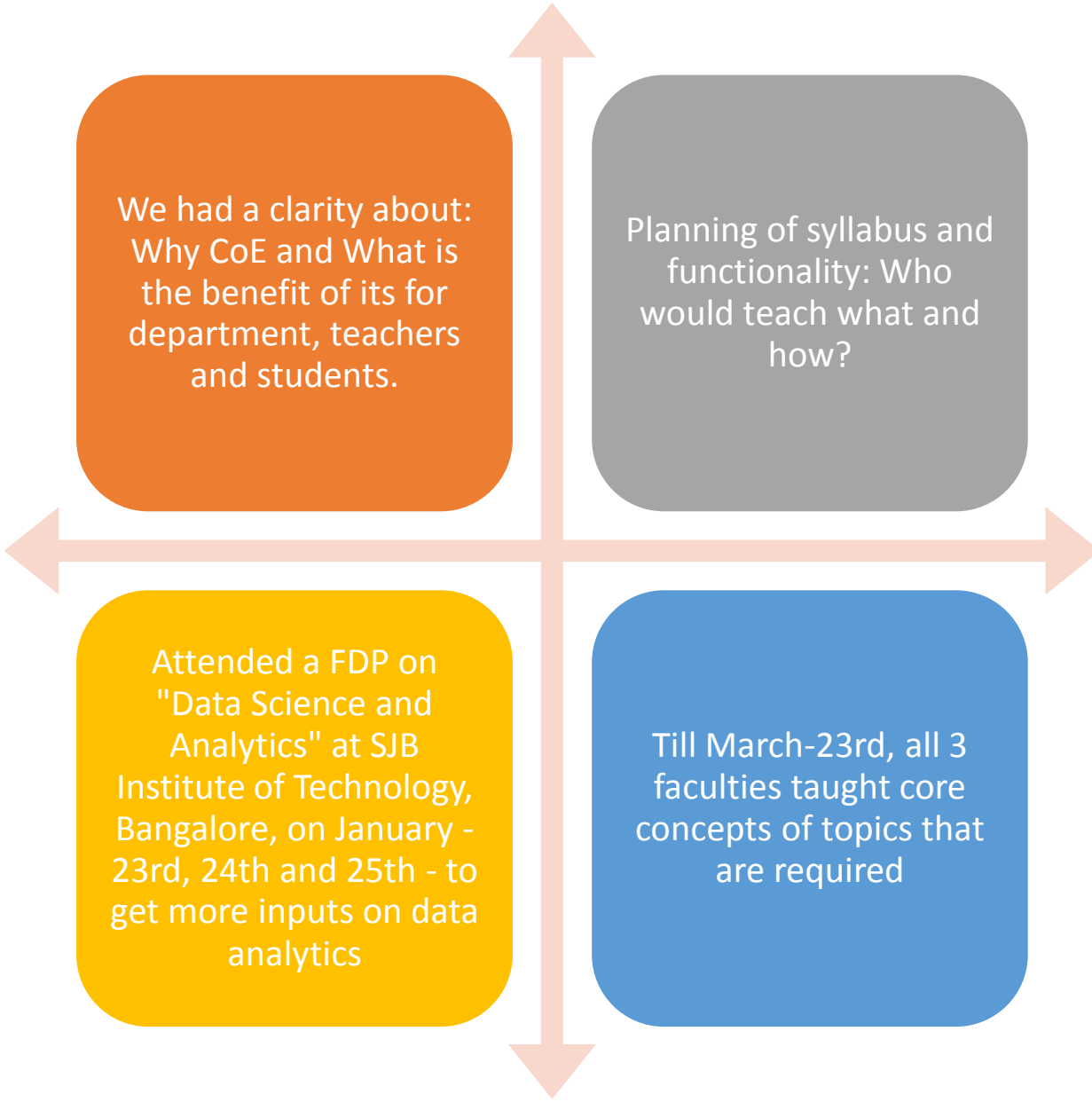


31-Jan-2019, Thursday - first session, i.e. the next day of Information Security CoE's workshop

We encouraged 6th semester students first, to join CoE

Initially, with 4 students we started: Monica, Namitha, Sushmitha, Shiva Prasad (all 6th sem students)

# Phase – 1: Learning Phase



# Phase – 2: Evaluation Phase

This was a kind of transformation that was brought in coe in a rapid way. And the reason behind this was one college level competition, that is - "DATATHON'19"

From March 24th, all the students (14 students out of 16) registered for this event, that was held on "APRIL-5th" at PES UNIVERSITY

The event was regarding: collect the data, clean the data, process the data and draw useful insights

To this event our one of the advisor - Dr. Viraj sir, was a mentor

By taking valuable tips from Mr. Sufian, from EC CoE, and Dr. Shanthi Mahesh, HOD of ISE, and faculty coordinators and also, alumni, Dr. Prajit Dhar, Dutch, taught them some data analytics through webinar.

Now, students understand the intensity at which they need to see data analytics on whole

Students are also been assigned with certain projects, which would be showcased in an exhibition along with final year student's projects.

Students will also be participating in "HACKATHON" by CSI.

# Strategic Plan for April to May- 2019

Students of CoE would take up ISE department result dataset and draw useful insights to improve the functionality and performance of the department. (In Summer Vacation at ATRIA Campus)

Agenda: To come up with a model that would give us useful conclusions when a dataset is fed as input.

CoE students would give hands-on sessions on data analytics and analysis to students who are part of other CoEs. (1 month)

Trying to connect with an institute and take-up their projects and build it in our CoE as consider it as internship work.


ANALYTICS
DATA ANALYTICS CENTER OF EXCELLENCE


DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING
ISE
73
83
69
Since 2001

# THANK YOU

# ALL SESSION REPORT

| Topic | Date |
|---|---|
| ▪ Introducing Big Data<br>▪ Why, how, where, the Big data is used? | 31/01/2019 |
| ▪ Kirk Borne's 10V's of Big Data<br>▪ Assignment-1 review<br>▪ Understanding the Big Data applications in several disciplines by defining them over 4 V's<br>▪ Large Hardon Collider (LHC), a real-time experiment example to understand the significance of Big Data | 04/04/2019 |
| ▪ Assignment-2 review<br>▪ Understanding the challenges related to Big Data<br>▪ The concept of distributed processing<br>▪ What is complex data and how its structure impacts during data processing<br>▪ A bottleneck problem in Big Data<br>▪ What is Hadoop? And how Hadoop solves the problems of Big Data<br>▪ HDFS: Hadoop Distributed File System, M-R: MapReduce<br>▪ A simple discussion on the working of, how MapReduce provides output to the wordcount problem | 05/02/2019 |
| ▪ HDFS Architecture and its working<br>▪ 5 services and daemons of HDFS<br>▪ Recap of previous session's topics discussion | 06/02/2019 |
| ▪ Basics of Python Programming<br>▪ Solving python programming quiz | 08/02/2019 |
| ▪ Installation of Cloudera-Hadoop in laptop | 09/02/2019 |
| ▪ Data mining concepts and techniques:<br>▪ Data analytics concepts<br>▪ Life cycle of data analytics<br>▪ Levels of Data Analytics | 11/02/2019 |
| ▪ Python programming practice | 12/02/2019 |
| ▪ Setting up single-node cluster in Hadoop | 13/02/2019 |
| ▪ Python programming: User defined functions, looping conditions in python | 15/02/2019 |
| ▪ Data mining techniques which includes classification<br>▪ Decision tree algorithms and numerical solving | 18/02/2019 |
| ▪ Recap session Big Data introduction | 19/02/2019 |

| | |
|---|---|
| ▪ Big Data challenges, How Hadoop can solve the challenges<br>▪ HDFS and MapReduce basics | |
| ▪ HDFS core components: Namenode, datanode and secondary namenode<br>▪ Secondary namenode: Checkpointing process<br>▪ HDFS data blocks<br>▪ HDFS write mechanism<br>▪ HDFS read mechanism<br>▪ MapReduce | 20/02/2019 |
| ▪ Data Structure operations in python<br>▪ File operators in Python | 22/02/2019 |
| ▪ Data Mining: Association Rule Analysis<br>▪ Apriori algorithm | 25/02/2019 |
| ▪ Data structure in python: Operations, methods and lists and functions | 26/02/2019 |

ATRIA INSTITUTE OF TECHNOLOGY
DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING
CENTER OF EXCELLENCE – DATA ANALYTICS

# FACULTY DETAILS

| | | |
|---|---|---|
|  | Name of the Faculty | Dr. Neha Mangla |
| | Designation | Associate Professor |
| | Department | Information Science and Engineering |

| | | |
|---|---|---|
|  | Name of the Faculty | Abhilash |
| | Designation | Assistant Professor |
| | Department | Information Science and Engineering |

| | | |
|---|---|---|
|  | Name of the Faculty | Ayesha Tarannum |
| | Designation | Assistant Professor |
| | Department | Information Science and Engineering |

ATRIA INSTITUTE OF TECHNOLOGY
DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING
CENTER OF EXCELLENCE – DATA ANALYTICS

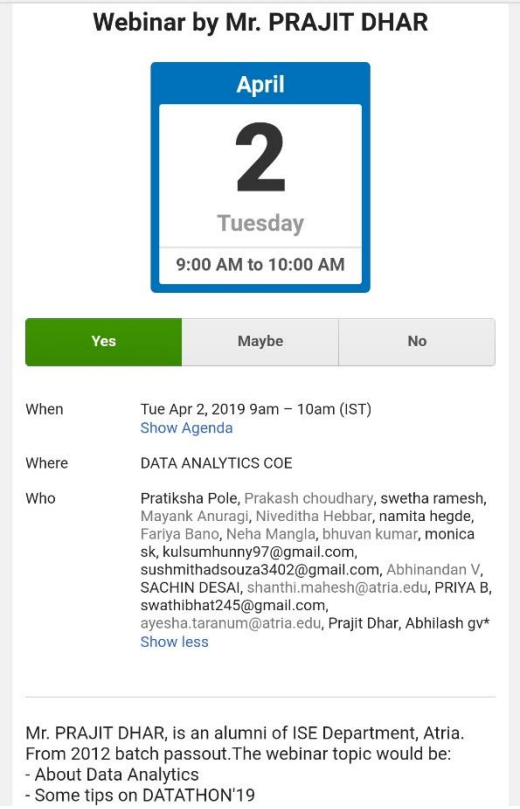
# STUDENT DETAILS

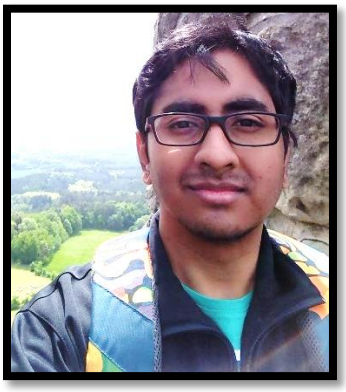
| SL.NO | STUDENT NAME | SEMESTER |
|---|---|---|
| 1 | Monica S Kumar | 7th |
| 2 | Namitha Hegde | 7th |
| 3 | Sushmitha | 7th |
| 4 | Priya B | 7th |
| 5 | Swathi K Bhat | 7th |
| 6 | Sachin | 7th |
| 7 | Kulsum | 7th |
| 8 | Mahima Krishna | 7th |
| 9 | Swetha Ramesh | 7th |
| 10 | Fariya Banu | 7th |
| 11 | Prakash Choudary | 7th |
| 12 | Prathiksha Pole | 7th |
| 13 | Niveditha | 7th |
| 14 | Simran | 7th |
| 15 | Abhinandan | 7th |
| 16 | Mayank | 5th |
| 17 | Neha Prakash | 5th |
| 18 | Ganesh A | Recently Passed Out |
| 19 | Narayana | Recently Passed Out |

# ATRIA INSTITUE OF TECHNOLOGY
# DEPARTMENT OF INFORMATION SCEINCE AND ENGINEERING
# CENTRE OF EXCELLENCE – DATA ANALYITCS

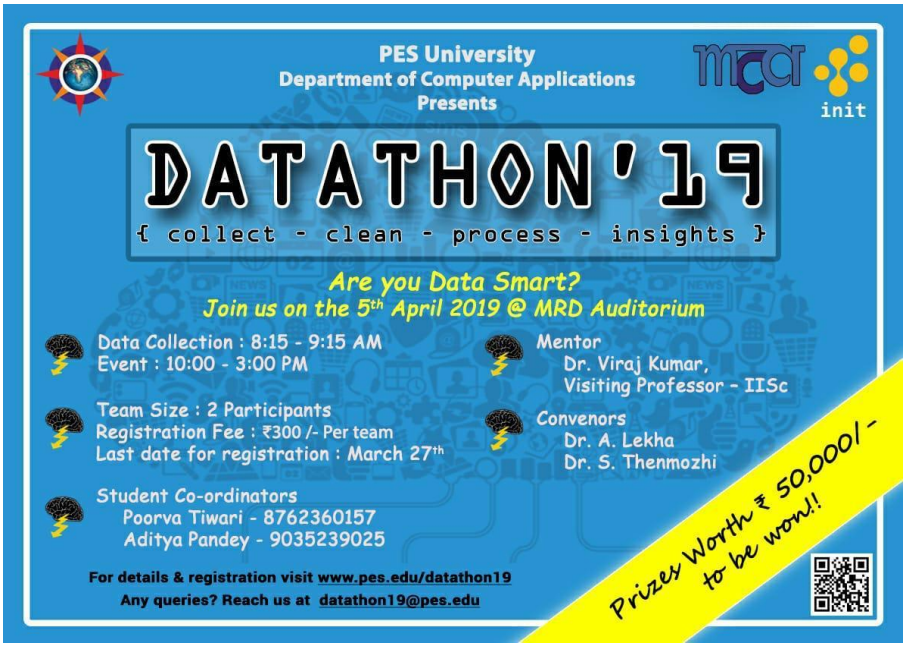
PES University, Main campus had conducted a 6-hour technical event – **"DATATHON'19: collect – clean – process – insights"** on 5th April 2019 from 8:15 a.m. to 5:00 p.m. The aim of this data event was to create awareness in finding useful data insights by generating and visualizing patterns and applying machine learning and data mining-based algorithms. In this event, around 12 students participated from the 4th, 6th and 8th semesters of the Department of Information Science and Engineering and got an idea on how to collect, pre-process, process and visualize the data.

In order to help students in preparing for this DATATHON'19 competition, there was a webinar conducted under CoE-Data Analytics, by one of the alumni of the ISE department, **MR. PRAJIT DHAR,** Former Research assistant at Fraunhofer IPA. The title of the webinar was **"DATA ANALYTICS"** and some tips on the competition. This webinar went on for 2 hours on **2nd April 2019** from 9 a.m. to 11 a.m. For this webinar around 16 to18 students took part and interacted with the speaker.



**MR. PRAJIT DHAR,**
Former Research assistant,
Fraunhofer IPA,
Germany


PES University
Department of Computer Applications
Presents
mcor
init
DATATHON'19
{ collect - clean - process - insights }
Are you Data Smart?
Join us on the 5th April 2019 @ MRD Auditorium
Data Collection : 8:15 - 9:15 AM
Event : 10:00 - 3:00 PM
Team Size : 2 Participants
Registration Fee : ₹300 /- Per team
Last date for registration : March 27th
Student Co-ordinators
Poorva Tiwari - 8762360157
Aditya Pandey - 9035239025
Mentor
Dr. Viraj Kumar,
Visiting Professor - IISc
Convenors
Dr. A. Lekha
Dr. S. Thenmozhi
Prizes Worth ₹ 50,000/- to be won!!
For details & registration visit www.pes.edu/datathon19
Any queries? Reach us at datathon19@pes.edu