

---

# Introduction to Data Science


## Course Structure (CS2004)

**Dr. Kusum Kumari Bharti**

Computer Science and Engineering Department,  
PDPM-Indian Institute of Information Technology Design and  
Manufacturing, Jabalpur  
Dumna Airport Road - 482005  
Email: [kusum@iiitdmj.ac.in](mailto:kusum@iiitdmj.ac.in)

---

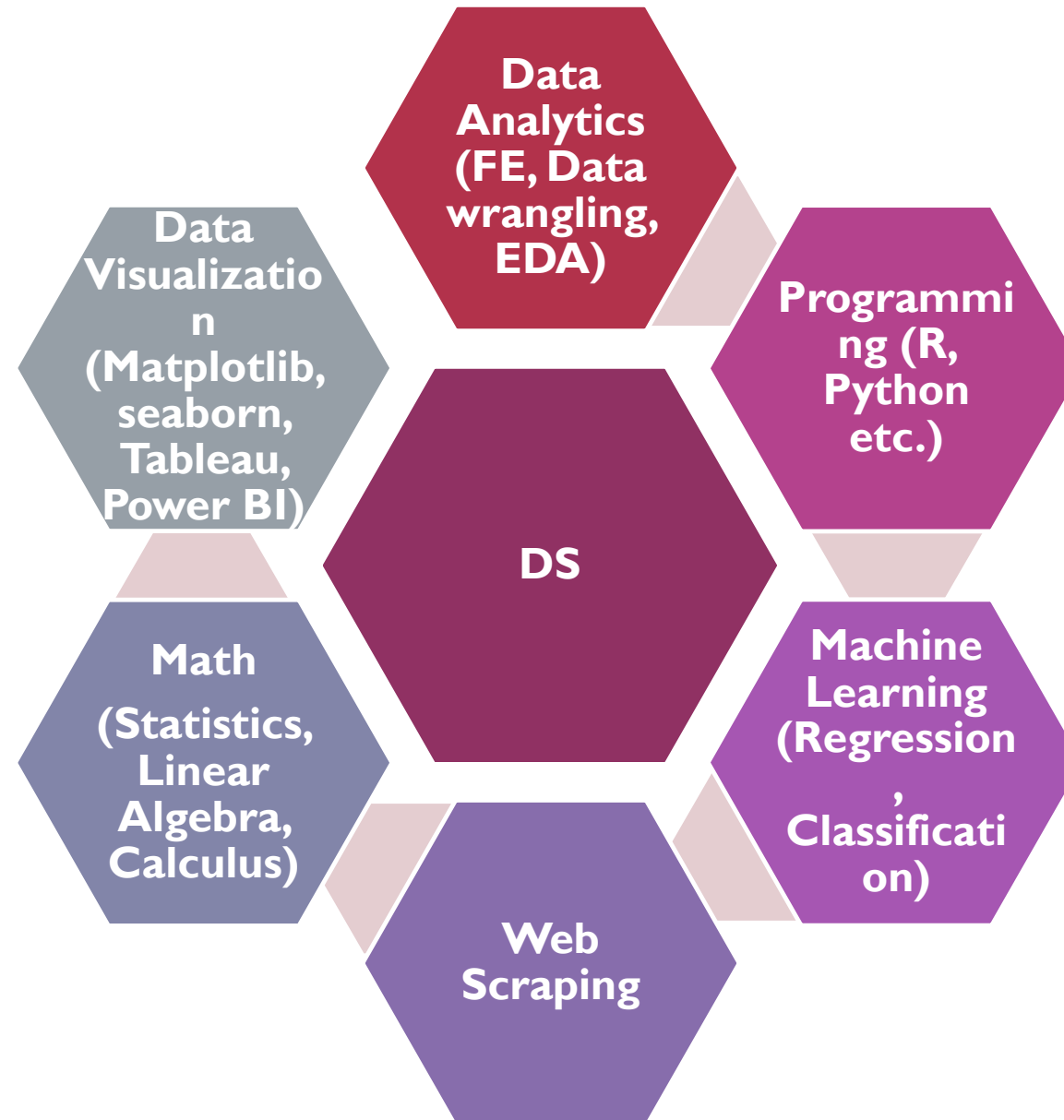
# Content

- Course Objective 
- Course Learning Outcome
- Course Home Page
- Lecture Plan
- Evaluation Scheme
- Self Learning Resources

---


# Course Objective

To elaborate the basics of data science and provide a foundation for understanding the challenges and applications.



---

# Content

- Course Objective
- Course Learning Outcome 
- Course Home Page
- Lecture Plan
- Evaluation Scheme
- Self Learning Resources

---


## COURSE LEARNING OUTCOMES (CLOS)

On completion of this course, the students will be able to:

- ❑ To analyse the need and usage of analytics and visualization techniques.
- ❑ To implement how to manage, manipulate, cleanse and analyse data.
- ❑ To Implement various data manipulation approaches in Python.
- ❑ To develop different type of data dashboard for data analysis using different tools.

---

# Content

- Course Objective
- Course Learning Outcome
- Course Home Page 
- Lecture Plan
- Evaluation Scheme
- Self Learning Resources

---

## Course Home Page (*Canvas*)


- **Course Home page consist of:-**

- Notifications related to: Exams, Quizzes, Assignment, Project work, Lab evaluation, submission deadlines, etc
- Slides
- Lab Experiments
- Assignments and Projects
- Books
- Supplement Resources



---

# Content

- Course Objective
- Course Learning Outcome
- Course Home Page
- Lecture Plan 
- Evaluation Scheme
- Self Learning Resources

---

# Lecture Plan

- Lectures per week = 3 (1hr)
- Lab per week = 2 (2hr)

# Evaluation Scheme

## Evaluation Scheme\*

Quizzes	:	20 (Quiz 1 02.03.2023 Quiz 2 04.10.2023)
Mid Term	:	25 (18.09.2023 to 23.09.2023)
End Term	:	35 (20.11.2023 to 25.11.2023)
Lab Exam	:	20 (Will be Announced in the Labs)

\* Tentative

---

# Lab plan

**Total Marks 20**

## **Tasks:**

- Lab assignment **Marks 10**
- Group based project on Real life applications – **Batch size 03 and Marks 10**

---

# Course plan

- **Part I (Aug 2023)** → Introduction to Data Science, Data, Basic statistics, Intermediate statistics
- **Part II (Sep 2023)** → Intermediate statistics, Advance Statistics, Exploratory data analysis
- **Part III (Oct 2023)** → Feature Engineering, Database
- **Part IV (Nov 2023)** → Machine Learning

---

# Lab plan

- **Part I (Aug 2023) → Crash course on Python**
  - Basic syntax, Advance of python
- **Part II (Sep 2023) → Cover Basics of Data Science**
  - Exploratory data analysis
- **Part III (Oct 2023) → Covers Basics of Data Science**
  - Exploratory data analysis, Databases
- **Part IV (Nov 2023) → Advances in Data Science**
  - Machine learning for Data Science (Data collection and Experimentation)

---

# Content

- Course Objective
- Course Learning Outcome
- Course Home Page
- Lecture Plan
- Evaluation Scheme
- Self Learning Resources



---

# Self Learning Resources

- Coursera
- Edx
- Udemy
- Datacamp
- Udacity
- MITOpenCourseware
- Kaggle
- Codecademy
- Alison
- Treehouse
- FutureLearn
- Harvard Extension School



---

# Self Learning Resources

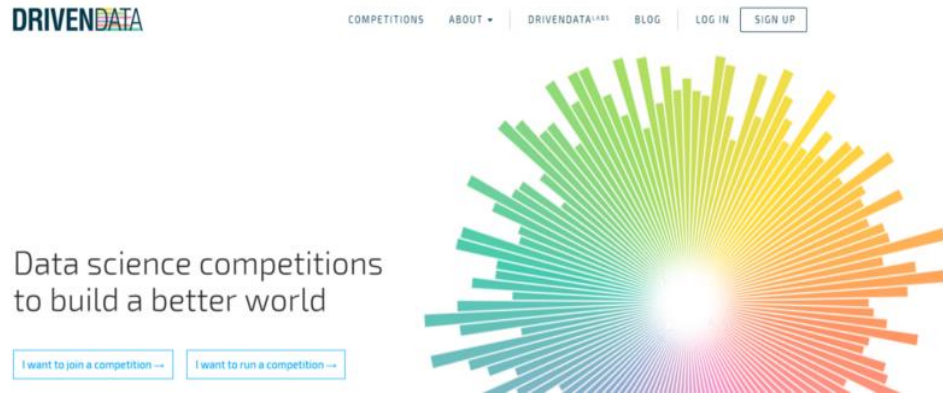
- Papers with codes (<https://paperswithcode.com/> )
- Projects (<https://github.com/trending> )
- Data Science news  
(<https://economictimes.indiatimes.com/topic/Analytics-and-Data-Science> )

---

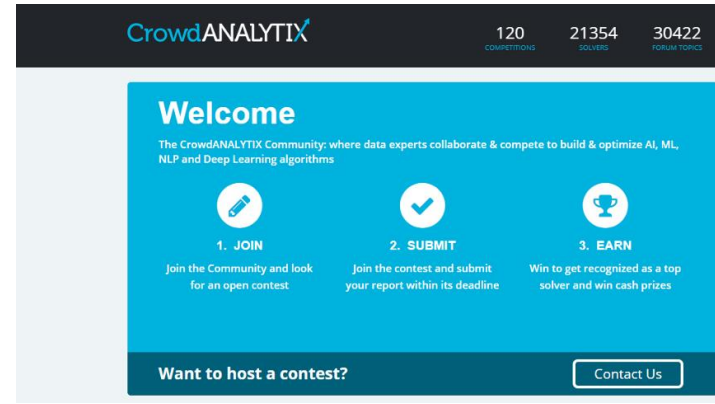
## Join and Explore LinkedIn

- Explore internships
- Explore freelancing projects
- Explore current trends in the job market
- Connect to your seniors

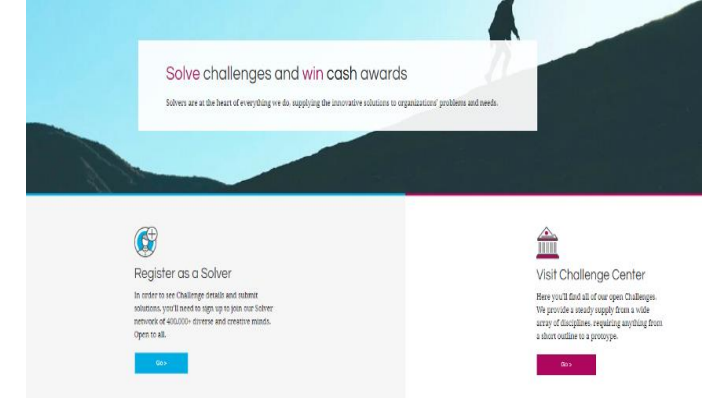
# TOP COMPETITIVE DATA SCIENCE PLATFORMS OTHER THAN KAGGLE



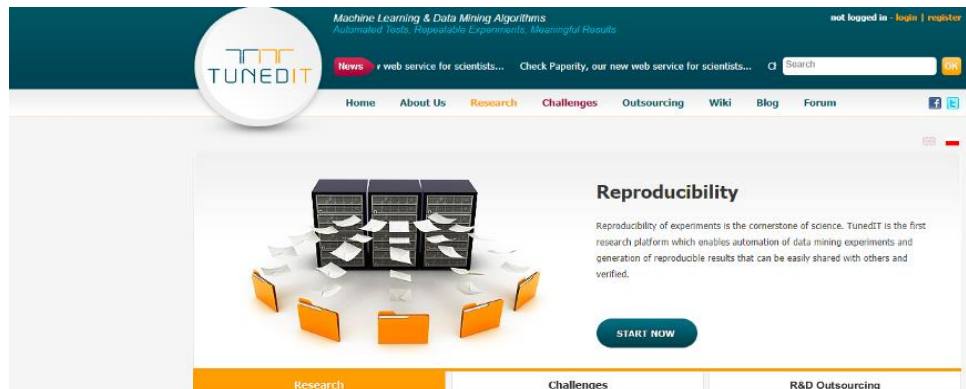
**Drivendata** hosts data science competitions for **social good** in areas like international development, health, education, research and conservation, and public services.



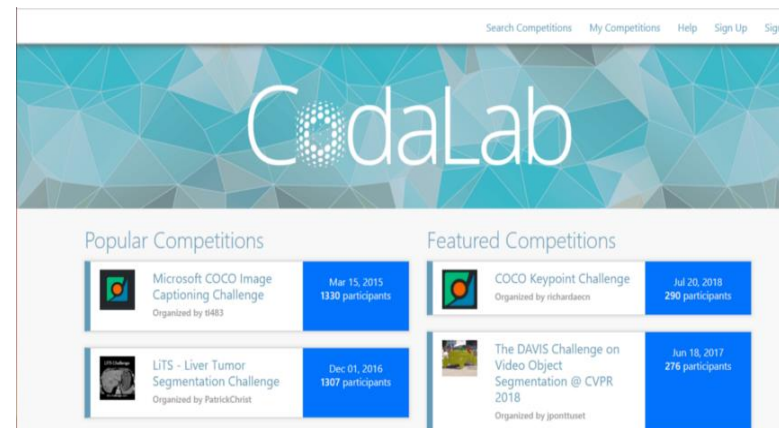
**CrowdANALYTIX** is a crowdsourced analytics platform that converts business challenges and problems into competitions.



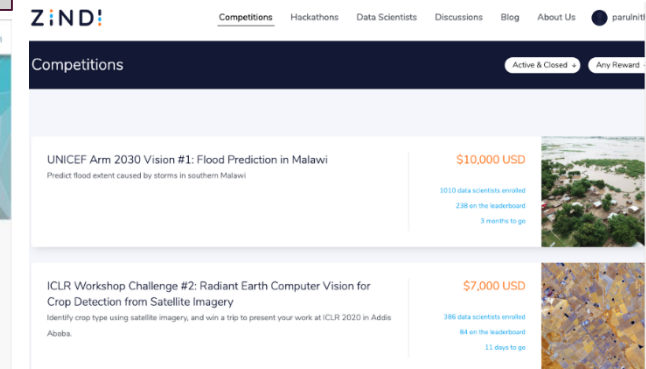
**InnoCentive** mainly focuses on problems dealing with life sciences



**TunedIT Challenges** platform for hosting data competitions — for educational, scientific and business purposes.



**Codalab** enables researchers, developers, and data scientists to collaborate, with the goal of advancing research fields



**Zindi** focused on solving Africa's most pressing problems.

## Welcome to Tianchi

Big data and distributed computing resources  
Cutting-edge solutions for real-world applications

Active

Algorithm

Innovative

Program

Getting Started

Dataset

### IJCAI-19 Alibaba Adversarial AI Challenge

Algorithm

Abstract: Alibaba (including Taobao and Tmall) is the world's largest e-commerce platform, which provides high-quality services to billions of customers. As an important role...

Rewards

\$39000

Teams

2376

Deadline of Season 1

2019-04-29

Sponsors Alibaba Security · IJCAI 2019 · Tianchi

Active

## Challenges

All 12 Active 14 Completed 38



LifeCLEF 2019 Plant  
Image-based plant identification on Amazonian flora

2490 Views 128 Participants 0 Submissions



ImageCLEF 2019 VQA-Med

23 days left

2422 Views 94 Participants 0 Submissions



ImageCLEF 2019 Tuberculosis - Severity scoring

The data science challenge platform [crowdAI](#) hosts multiple open data science challenges each year.

Data scientists cooperate with each other and connect with businesses and governments globally to solve the hardest business problems across industries.

## CONTESTS

Switch to events

Students only

All Contests

ML Hack

Skill Tests

Case study

Visualization

Blogathon

Excel

Hiring

Practice



LTFS Data Science FinHack (ML Hackathon)

13-04-2019 to 21-04-2019

Cash Prizes of INR 3.5 Lacs along with Job Opportunities with LTFS

2468 registered

162 teams

PARTICIPATE



Club Mahindra DataOlympics

03-05-2019 to 05-05-2019

Prizes worth 2 lacs along with Job Opportunities with Club Mahindra

81 registered

Free

PARTICIPATE



ML Hikeathon (An Online Machine Learning Hackathon)

30-03-2019 to 07-04-2019

Prizes worth INR 3.5 lacs along with Job Opportunities with Hike

5374 registered

496 teams

PARTICIPATE

## DATA SCIENCE CHALLENGE

### PAST CHALLENGES

Multi label classification

#### GROWING INSTABILITY: CLASSIFYING CRISIS REPORTS

Winner: qingschenwang

SEE THE RESULTS >



£20,000



579

Image analysis: Detect and Classify

#### SAFE PASSAGE: DETECTING AND CLASSIFYING VEHICLES IN AERIAL IMAGERY

Winner: gbarbadillo

SEE THE RESULTS >



£20,000



717



HOME

PROGRAM

ATTENDING

NEWS & CALLS

KDD CUP

SPONSORS

ORGANIZING COMMITTEE

## KDD Cup 2019

Home / KDD Cup

KDD Cup is the annual Data Mining and Knowledge Discovery competition organized by ACM Special Interest Group on Knowledge Discovery and Data Mining, the leading professional organization of data miners. SIGKDD-2019 will take place in Anchorage, Alaska, US from August 4 - 8, 2019. The KDD Cup competition is anticipated to last for 2-4 months, and the winners will be notified by mid-July 2019. The winners will be honored at the KDD conference opening ceremony and will present their solutions at the KDD Cup workshop during the conference.

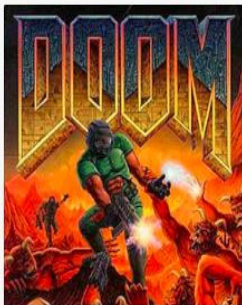
This year, there are three competition tracks:

- Regular Machine Learning Competition Track (Regular ML Track)
- Automated Machine Learning Competition Track (Auto-ML Track)
- "Research for Humanity" Reinforcement Learning Competition Track (Humanity RL Track)

**KDD Cup** is the annual Data Mining and Knowledge Discovery competition

**Analytics Vidhya** provides a community-based knowledge portal for Analytics and Data Science professionals. Analytics Vidhya also offer job opportunities to the top scorers.

The challenges are designed to encourage the brightest minds in data science to help solve real-world problems.



## Visual Doom AI Competition 2018 - Singleplayer Track (1)

ViZDoom Reinforcement Learning



By Poznan University of Technology

Completed 11684 Views 65 Participants 204 Submissions



45

FOLLOW

[Overview](#) [Leaderboard](#) [Discussion](#) [Dataset](#) [Winners](#)

ViZDoom is a **Doom**-based **AI** Research Platform for Reinforcement Learning from Raw **Visual** Information.

## Machine Learning Contests

Discover ongoing machine learning competitions/data science contests across Kaggle, DrivenData, AICrowd, and other platforms.

Title	Type	Deadline	Prize Pool	Platform	Sponsor
<a href="#">March Madness: Basketball Predictions (Women's)</a>	Supervised Learning	20 Mar 2020	\$25,000	Kaggle	Google Cloud
<a href="#">March Madness: Basketball Predictions (Men's)</a>	Supervised Learning	20 Mar 2020	\$25,000	Kaggle	Google Cloud
<a href="#">Predict the DEFCON level</a>	Supervised Learning	20 Mar 2020	\$475	HackerEarth	HackerEarth
<a href="#">Classify Disaster Tweets</a>	NLP	23 Mar 2020	\$10,000	Kaggle	Google Cloud
<a href="#">Calculate Emissions Factors For Electricity</a>	Analysis/Visualisation	24 Mar 2020	\$25,000	Kaggle	Google
<a href="#">Classifying Crop Types</a>	Supervised Learning	29 Mar 2020	\$5,000	Zindi	Radiant Earth/ICLR
<a href="#">Detecting Crop Diseases</a>	Computer Vision	29 Mar 2020	\$5,000	Zindi	CGIAR/ICLR

Machine Learning Contests is a data science competition aggregator site.

Source: <https://towardsdatascience.com/top-competitive-data-science-platforms-other-than-kaggle-2995e9dad93c>



---

# Finally



## Learning by Doing

---

*Thank  
you*

