Introduction to Data Science

Course Structure (CS2004)

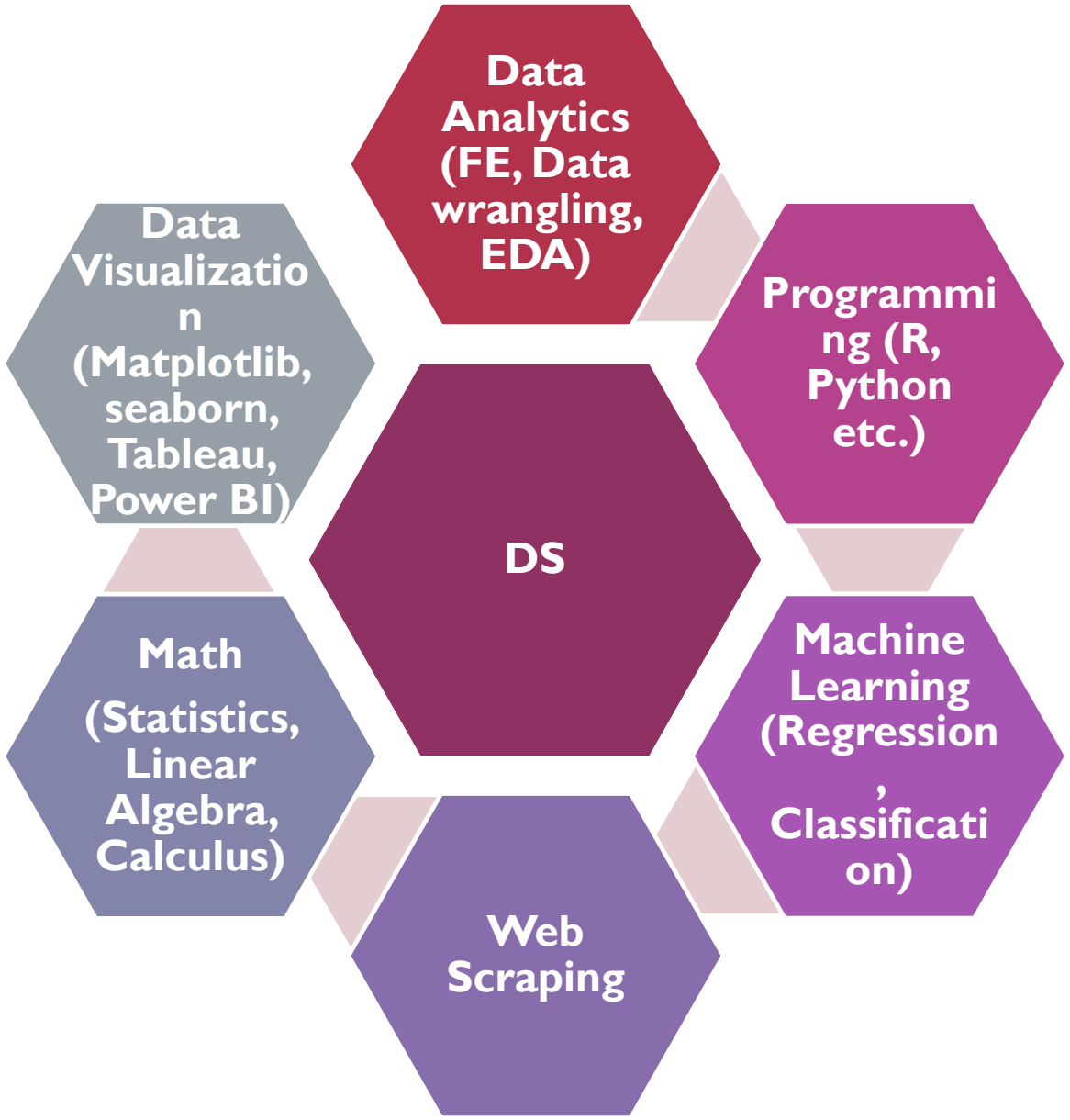
Dr. Kusum Kumari Bharti
Computer Science and Engineering Department,
PDPM-Indian Institute of Information Technology Design and Manufacturing, Jabalpur
Dumna Airport Road - 482005
Email: kusum@iiitdmj.ac.in

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.


Data Analytics (FE, Data wrangling, EDA)
Programming (R, Python etc.)
Machine Learning (Regression, Classification)
Web Scraping
Math (Statistics, Linear Algebra, Calculus)
Data Visualization (Matplotlib, seaborn, Tableau, Power BI)
DS

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
- Hardvard Extension School

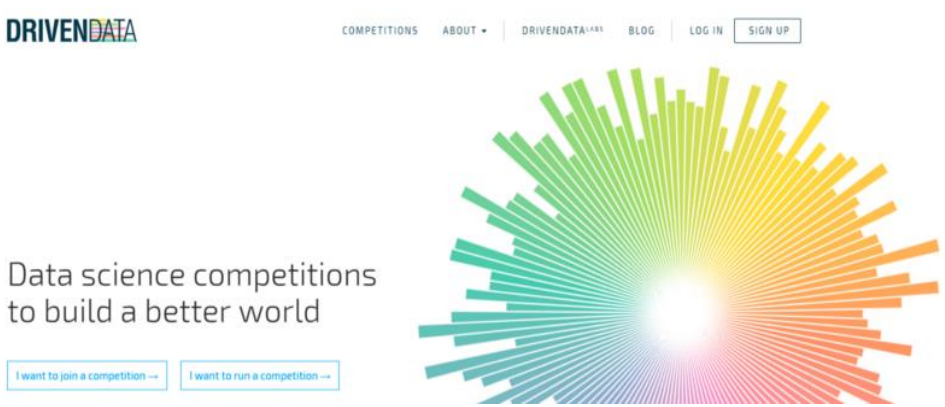
Self Learning Resources

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

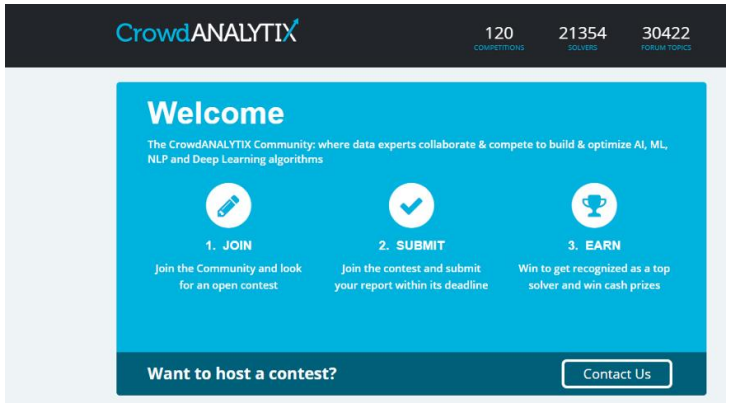
Join and Explore Linkedin

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

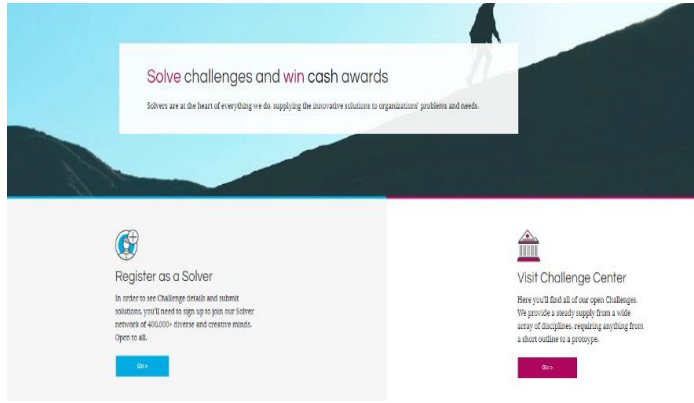
TOP COMPETITIVE DATA SCIENCE PLATFORMS OTHER THAN KAGGLE



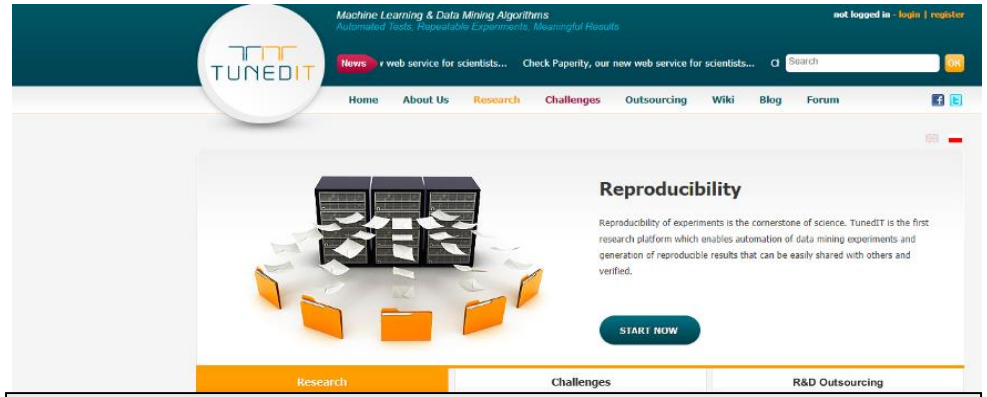
Drivendata hosts data science competitions for **social good** in areas like international development, health, education, research and conservation, and public services.



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

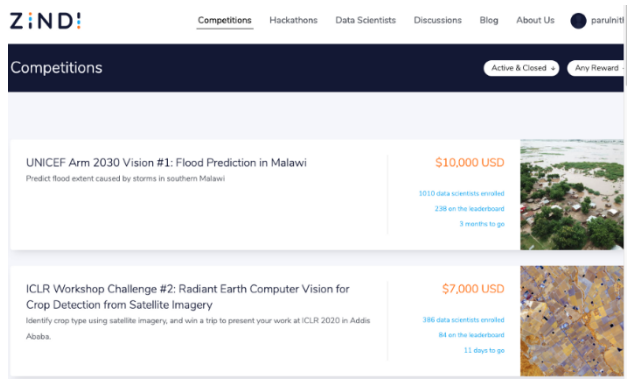


InnoCentive mainly focuses on problems dealing with life sciences

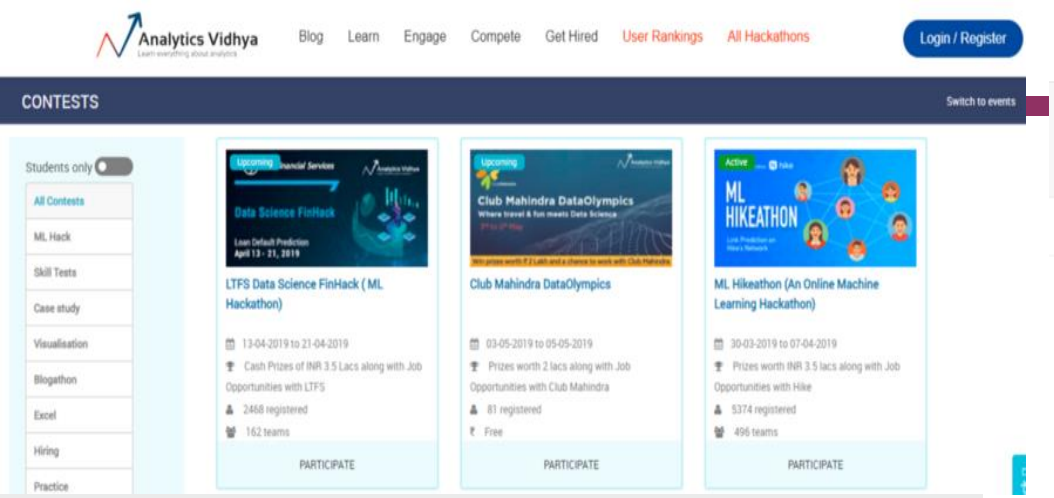


TunedIT Challenges platform for hosting data competitions — for educational, scientific and business purposes.

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

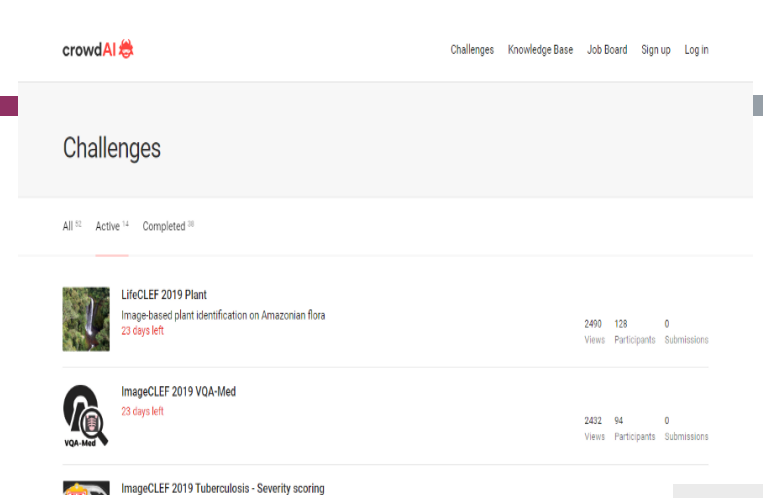


Zindi focused on solving Africa's most pressing problems.

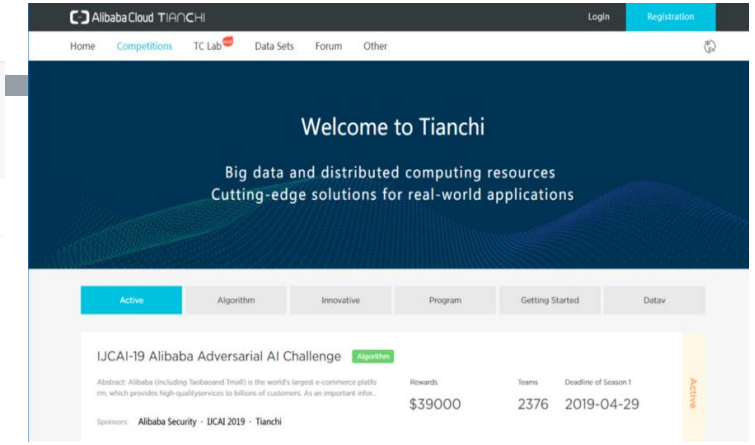


[Analytics Vidhya](#) provides a community-based knowledge portal for Analytics and Data Science professionals.

Analytics Vidhya also offer job opportunities to the top scorers.

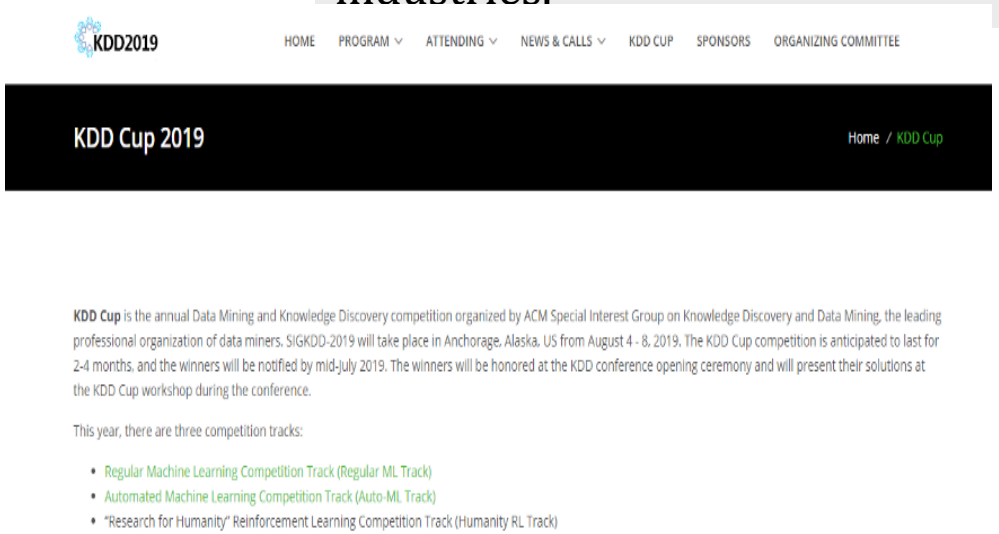


The data science challenge platform [crowdAI](#) hosts multiple open data science challenges each year.

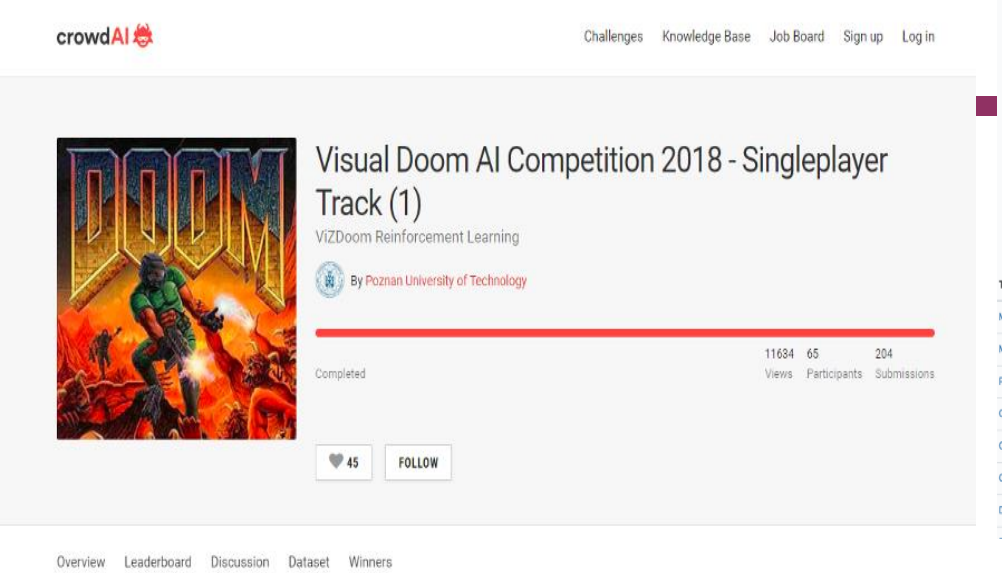


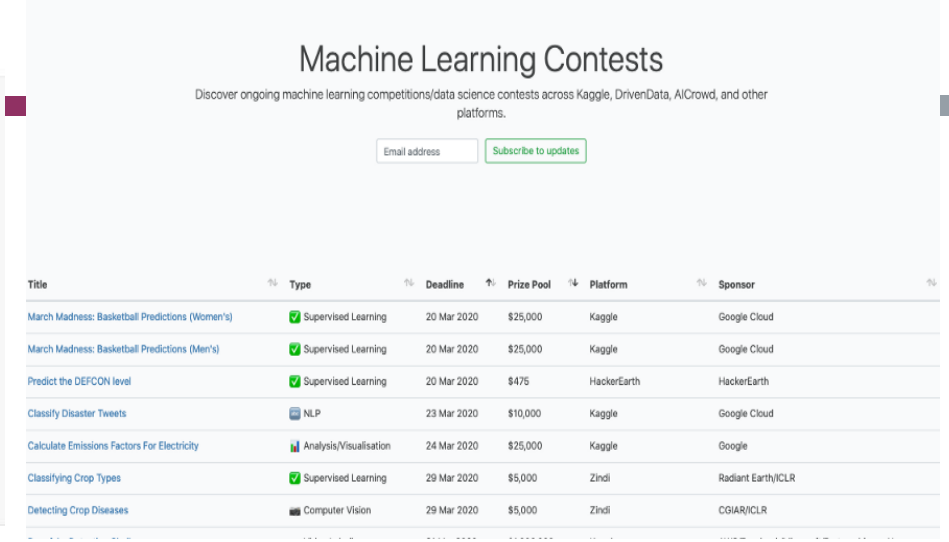
Data scientists cooperate with each other and connect with businesses and governments globally to solve the hardest business problems across industries.

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



[KDD Cup](#) is the annual Data Mining and Knowledge Discovery competition





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

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