Kaggle
A platform for making data science a sport

Seminar on Internet Technologies (WS2014/15)

Supervised by : Narisu Tao

Presented by

Hari Raghavendar Rao Bandari(11334055)

The sexiest job of 21st Century

hbr.org

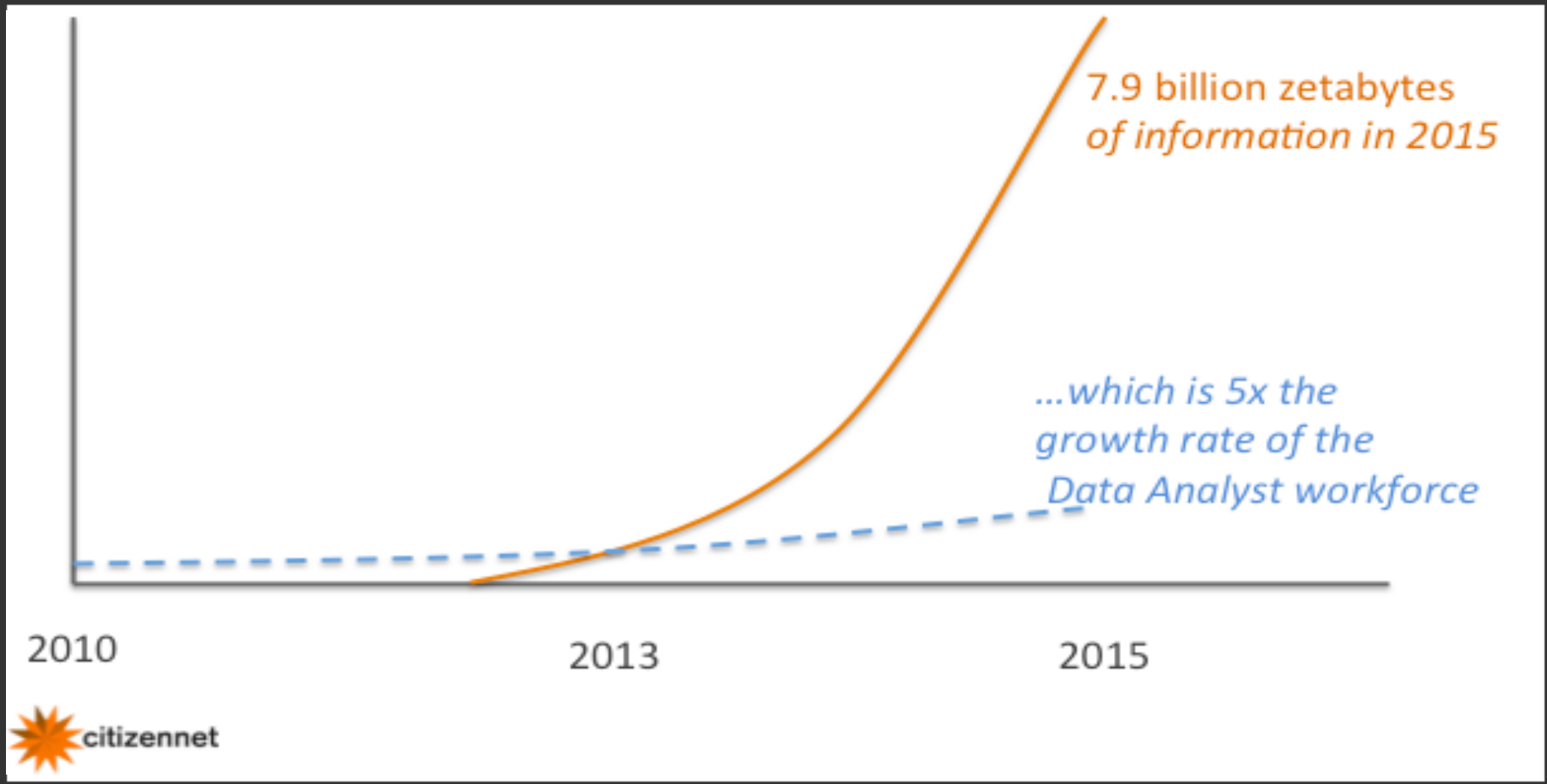
Harvard Business Review

Data Scientist:The Sexiest job of the 21st Century

by Thomas H. Davenport and D.J.Patil

Mckinsey estimates 140,000-190,000 shortage by 2018

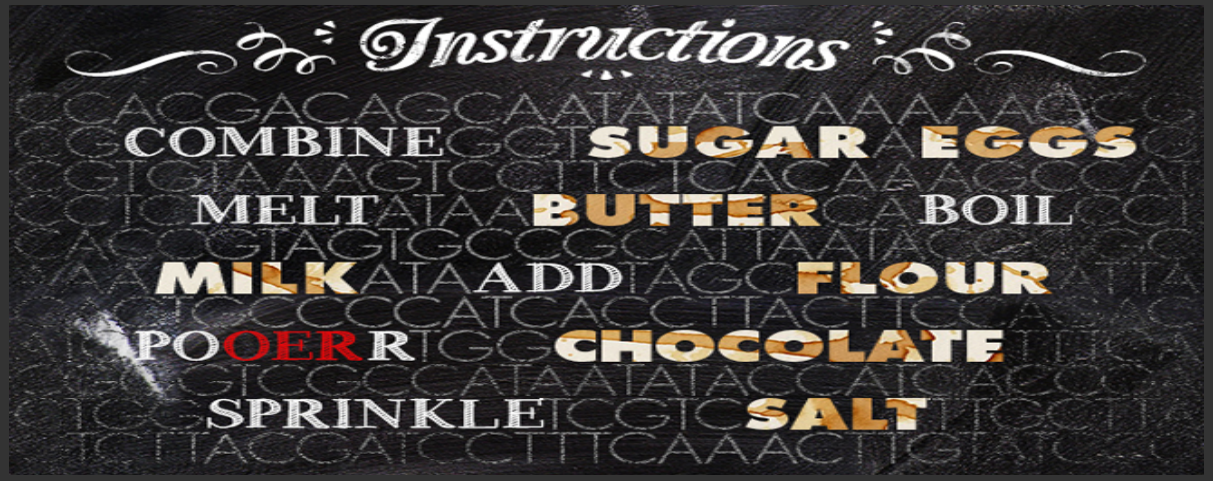
Why we are talking about "Data Science" currently



How Data Science Help To Stretch "Boundary of Human knowledge":

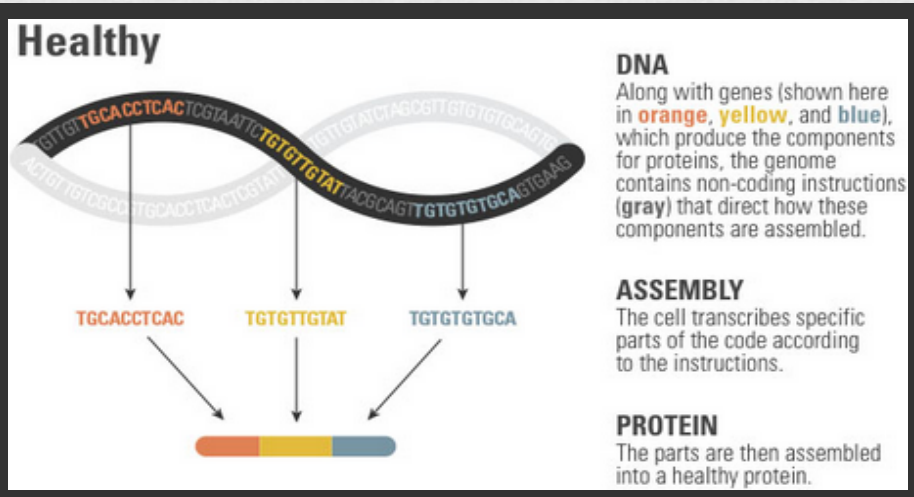
Machine Intelligence Cracks Genetic Contr

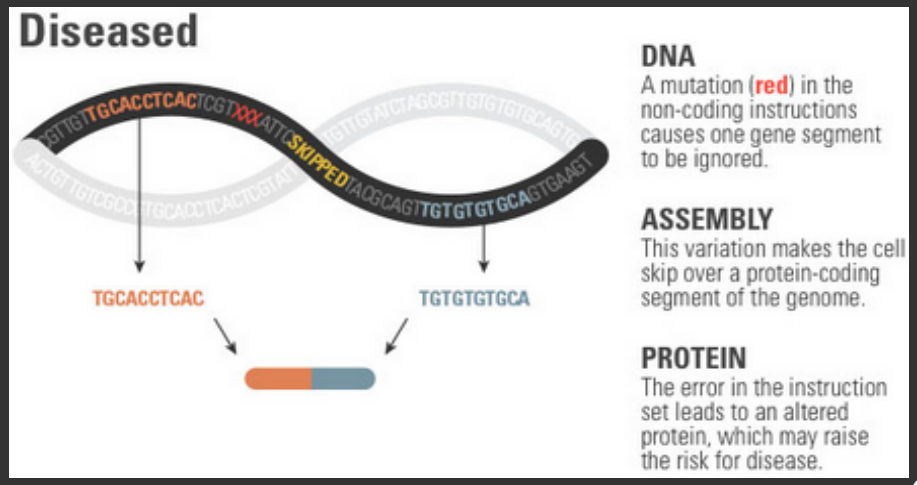
"small" part of genome which produce protein is considered but whole genome is important for Autism research



Eve th instructions and ingredients. So does the human genome. An error in the instructions can raise the risk for disease.

INSTRUCTIONS IN THE CODE



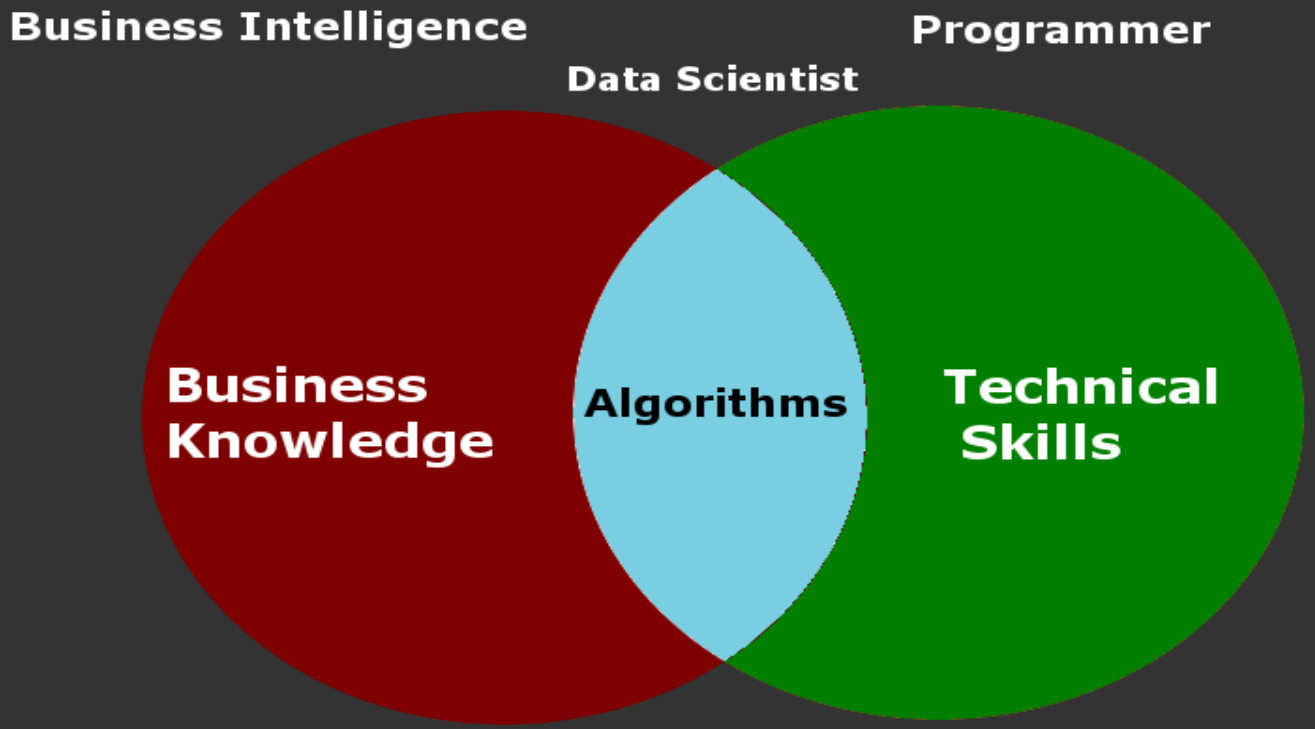


What kinds of skills are owned by Data scientist?

Skill are of two types:

1.Technical Skills

2.Business knowledge



Technical Skills:

Theoretical knowledge:

Mathematics & Statistics
linear algebra/R, SAS,Stata etc

Data Mining
Cluster analysis, anomaly detection, dependencies.

Data Modeling
UML class diagrams,CRC cards

Predictive Modeling
what the future holds

Machine Learning
develop and improve algorithms

Visualization
Google Visualization API

Practical skills:

Programming/ Scripting Languages
Python,java,Ruby,Perl,Pig

Distributed Computing Systems
Hadoop,Hbase,HIVE,Mapreduce,Cassandra

Relational Databases
SQL/NOSQL

Business Knowledge :

Domain Expertise

Know Organization data

Creativity and Curiosity

By trial and error methods

Storytelling

selling your discovery

Project Management

Budget, Time constraints, Profitable.

Ethics

Technology Policy

The Elevator Speech

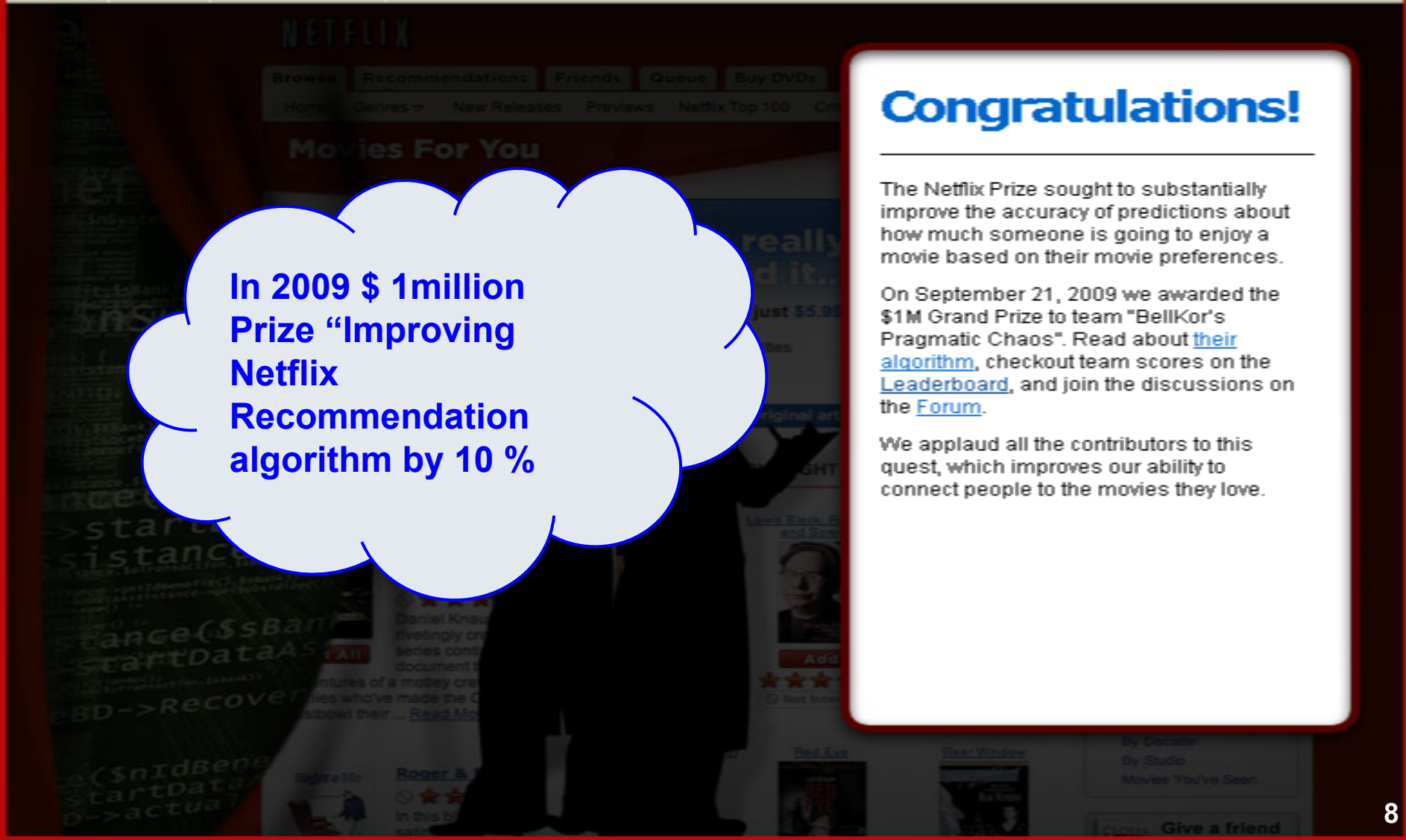
Simply define a profession

NETFLIX

Netflix Prize

COMPLETED

Home | Rules | Leaderboard | Update



In 2009 $ 1million Prize “Improving Netflix Recommendation algorithm by 10 %

Congratulations!

The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences.

On September 21, 2009 we awarded the $1M Grand Prize to team "BellKor's Pragmatic Chaos". Read about their algorithm, checkout team scores on the Leaderboard, and join the discussions on the Forum.

We applaud all the contributors to this quest, which improves our ability to connect people to the movies they love.

History of Kaggle:

Kaggle was founded by "Anthony Goldbloom" in 2010 in Melbourne,and moved to San Francisco in 2011. In November 2011, Kaggle announced Series A funding led by Index Ventures and Khosla Ventures.

Anthony Goldbloom
Founder and CEO

Ben Hamner
Cheif Data Scientist

Johnathan Hodge
VP Product

kaggle

Sign up Login

The Home of Data Science

COMPETITIONS - ENERGY INDUSTRY SPECIALIST
PREDICTIVE MODELING SOLUTIONS - TUTORI
KAGGLE IN CLASS - DATA SCIENCE JOBS BO

COMPETITIONS

HIRING DATA SCIENTIS

FOR CUSTOMERS

FOCUSED ON ENERGY

The Energy Industry uses our expertise in machine learning & big data to drive high-stakes decisions.

Find out more ›

COMPETITIVE ANALYTICS

For any industry, we tap the world's largest community of data scientists to solve your data problem.

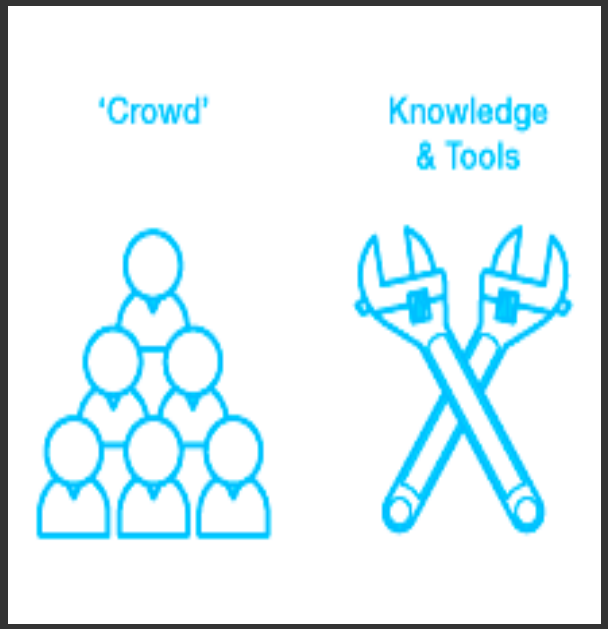
Host a competition ›

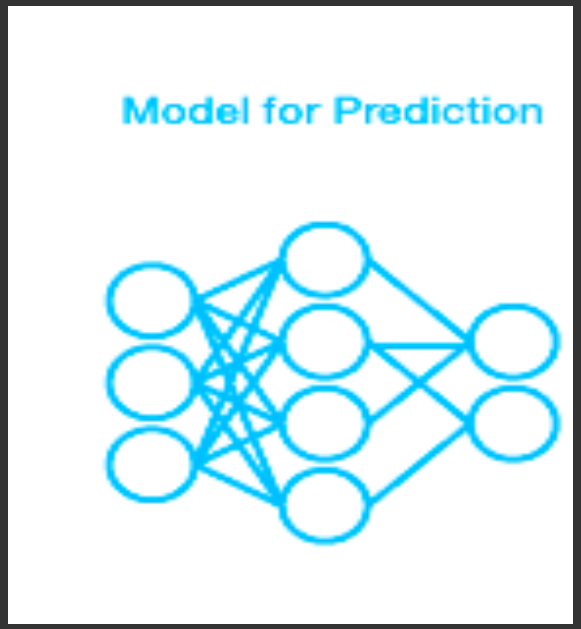
How Kaggle competitions work?

Competition Host
(Comapanies,University Groups,Researchers)

Data Scientist's participate in Competition's

Best Predictive Modeling "Win Competition"





Stage 1

Stage 2

Stage 3

kaggle

Customer Solutions | Competitions | Community

"astronomers studying for years"
"A Phd student from Cambridge university solved in 10 days."

Completed • $3,000 • 70 teams

Mapping Dark Matter

Mon 23 May 2011 – Thu 18 Aug 2011 (3 years ago)

Dashboard

- Home
- Data
- Information
  - Description
  - Background
  - Evaluation
  - Rules
  - Prizes
  - Ellipticity
  - FAQs

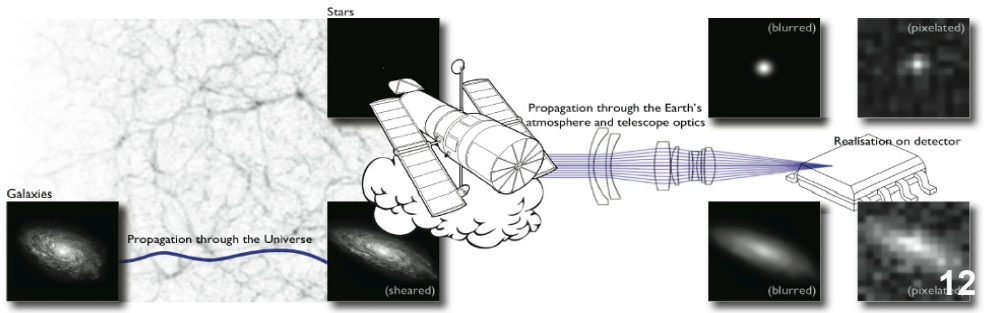
Supported by NASA and the Royal Astronomical Society. A cosmological image analysis competition to measure the small distortion in galaxy images caused by dark matter. The prize is an expenses paid visit to the NASA Jet Propulsion Laboratory (JPL).

Started: 1:42 pm, Monday 23 May 2011 UTC

Ended: 12:00 am, Thursday 18 August 2011 UTC (86 total days)

Points: this competition awarded 1.5X ranking points

Tiers: this competition counted towards tiers



kaggle

Sign up Login

The Home of Data Science

COMPETITIONS - ENERGY INDUSTRY SPECIALISTS
PREDICTIVE MODELING SOLUTIONS - TUTORIALS
KAGGLE IN CLASS - DATA SCIENCE JOBS BOARD

DATA SCIENTIST JOB BOARD

HIRING DATA SCIENTISTS? POST TO OUR JOBS BOARD ›

FOR CUSTOMERS

FOCUSED ON ENERGY

The Energy Industry uses our expertise in machine learning & big data to drive high-stakes decisions.

Find out more ›

COMPETITIVE ANALYTICS

For any industry, we tap the world's largest community of data scientists to solve your data problem.

Host a competition ›

User Ranking And Tier System:

Points

users are allocated points for their performance in competitions.

Tiers

NOVICE

New member on the site (not yet earned any ranking points)

KAGGLER

Members of the Kaggler tier who achieve 5000 points from competition performances have access to the private messaging system to contact other users on the site.

MASTER

Must fulfill 2 criteria:
Consistency: at least 2 Top 10% finishes in public competitions
Excellence: at least 1 of those finishes in the top 10 positions

How Kaggle change life of Data Scientist:

Jeremy Howard

Now CEO "Enlitic " (startedup in August 2014)

Past :President/Chief Scientist of "Kaggle"(top ranked in 2010/11)

Before: Management consulting McKinsey & Company and AT Kearney

Owen Zhang

Now chief Product officer at "DataRobot"

Past:(American international grp)

"Kaggle" 1st Rank (774,329 pts) and 28 competitions (2013-15)

Before: works at Travelers Insurance

Summary:

- Data Scientists are important in 21st century.
- kaggle host Competitions to solve data science problem.
- kaggle has changed the life of Data Scientist.

Thank you

Any Questions?

References:

Harvard Business Review

https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/

Mckinsey & company

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

http://blog.pivotal.io/pivotal/news-2/mckinsey-report-highlights-the-impending-data-scientist-shortage

Citizennet.com

http://citizennet.com/blog/2014/01/03/four-trends-in-marketing-technology-for-2014/

Netflix

http://www.netflixprize.com/index

Mastersindatascience.org

http://www.mastersindatascience.org/blog/data-scientist-foundations-the-hard-and-human-skills-you-need/

WIRED.com

http://www.wired.com/2014/12/machine-intelligence-cracks-genetic-controls/