Machine Learning
25 Best Jobs in America for 2015

Glassdoor Team | January 27, 2015

Whether you want a new job, or just want to make sure you already have a great job, Glassdoor has released its inaugural report highlighting the 25 Best Jobs In America for 2015.

This new report identifies the 25 best jobs based on each job’s overall Glassdoor Job Score*. The Glassdoor Job Score is determined using three key factors – earning potential based on average annual base salary, career opportunities rating and number of job openings. The jobs that made this list stand out across all three categories. Check out the complete results:

1. Physician Assistant – Glassdoor Job Score: 4.8
   - Number of Job Openings: 45,484
   - Average Base Salary: $111,376
   - Career Opportunities Rating: 3.5

2. Software Engineer – Glassdoor Job Score: 4.6
   - Number of Job Openings: 104,828
   - Average Base Salary: $98,074
   - Career Opportunities Rating: 3.3

3. Business Development Manager – Glassdoor Job Score: 4.6
   - Number of Job Openings: 11,616

* Glassdoor Job Score is determined using data from May 2014 – March 2015.
8. **Product Manager** - Glassdoor Job Score: 4.5
   - Number of Job Openings: 10,294
   - Average Base Salary: $113,363
   - Career Opportunities Rating: 3.3

9. **Data Scientist** - Glassdoor Job Score: 4.4
   - Number of Job Openings: 3,449
   - Average Base Salary: $104,476
   - Career Opportunities Rating: 3.8

10. **Sales Manager** - Glassdoor Job Score: 4.4
    - Number of Job Openings: 26,193
    - Average Base Salary: $76,556
    - Career Opportunities Rating: 3.3

11. **Solutions Architect** - Glassdoor Job Score: 4.4
    - Number of Job Openings: 3,028
A new species of techie is in demand these days—not only in Silicon Valley, but also in company headquarters around the world. “Data scientists are the new superheroes,” says Pascal Clement, the head of Amadeus Travel Intelligence in Madrid. The description isn’t exactly hyperbolic: The qualifications for the job include the strength to tunnel through mountains of information and the vision to discern patterns where
Data Science is the extraction of knowledge from large volumes of data that are structured or unstructured, which is a continuation of the field data mining and predictive analytics, also known as knowledge discovery and data mining (KDD).

Data science - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Data_science

IBM - What is a Data Scientist? - Bringing big data to the... www.ibm.com/software/datainfoosphere/data-scientist/ IBM - About data scientists. Rising alongside the relatively new technology of big data is the new job title data scientist. While not tied exclusively to big data projects, the data scientist role does complement them because of the increased breadth and depth of data being examined, as compared to traditional roles.

8 Skills You Need to Be a Data Scientist - Udacity - Climb ... blog.udacity.com/2014/11/data-science-job-skills.html Udacity - Nov 7, 2014 - Many resources out there may lead you to believe that becoming a data scientist requires comprehensive mastery of a number of fields, such as software development, data mining, databases, statistics, machine learning and data visualization. Don't worry. In my experience as a data scientist, that's not the case.

How do I become a data scientist? An evaluation of 3 ... datacopanalytics.com/.../how-do-i-become-a-data-scientist-an-evalu... - Aug 4, 2014 - One of the most frequent questions we hear, right behind "so, what exactly is a data scientist" or "what makes a great data scientist", is "how do i ...

Data science - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Data_science

In the news

Luring The Curious Into The Data Science Kitchen
Forbes - 2 days ago
Data scientists are like the master chefs of the business world. Sequestered in their glass ...

How Data Science Shaped This Teen\'s Counseling &amp; Text Service
Co.Exist - 22 hours ago

More news for what is a data scientist
Data science

From Wikipedia, the free encyclopedia

This article's tone or style may not reflect the encyclopedic tone used on Wikipedia. See Wikipedia's guide to writing better articles for suggestions. (February 2014)

Data Science is the extraction of knowledge from large volumes of data that are structured or unstructured,¹ Which is a continuation of the field data mining and predictive analytics, also known as knowledge discovery and data mining (KDD). "Unstructured data" can include emails, videos, photos, social media, and other user-generated content. Data Scientists are qualified people with strength and patience to tunnel through mountains of information and the technical skills in writing algorithms to extract insights from these troves of information.

Contents [hide]

1 Overview
2 History
3 Domain specific interests
4 Criticism
5 Research areas
  5.1 Security Data Science
  5.2 Clinical data science
  5.3 Genomic data science
6 Further reading
7 References

Overview [edit]

Data science employs techniques and theories drawn from many fields within the broad areas of nanotechnologies, physics, robotics, mathematics, statistics, information theory and information technology, including signal processing, probability models, machine learning, statistical learning, data mining, database, data engineering, pattern recognition and learning, visualization, predictive analytics, uncertainty modeling, data warehousing, data compression, computer programming, and high performance computing. Methods that scale to Big Data are of particular interest in data science, although
What is Data Science?

• Roughly defined: how one can use data to answer questions or guide decisions

• Research areas:
  • Computer Science
  • Machine learning
  • Data mining
  • Statistics
Machine Learning

• Learning: improving with experience at some task
  • Improve over task
  • with respect to some performance measure
  • based on some experience
• Writing computer programs that write computer programs
Rest of Video

- Three machine learning success stories
- One cautionary tale
- Machine learning research
Machine Learning Story 1
Face Detection & Recognition
Find photos by what’s in them

Looking for that photo of your pup? Just tap “dog” or the place you took it to find it faster.
What Does a Human Face Look Like?
if pixel153 > 128 & pixel154 > 128 & pixel155 > 128 & pixel156 < 64 & sqrt(pixel157) < 82 & log(pixel1132 * pixel1133) > 1 .... then image is a face*

* (not a real face recognition program)

Apple II image from wikipedia.com.
Eyes added digitally.
Machine Learning Story 2
Recommender Systems
Machine Learning Story 3
EMBERS: Predicting the News with Social Media Data
The EMBERS Project Can Predict the Future With Twitter

BY LEAH MCGRATH GOODMAN / MARCH 7, 2015 12:15 PM EST
Sindicato Unificado de Trabajadores se reunirá mañana!!! http://t.co/none
This is just beginning ... this has not been a democratic election

Today #tauro do not hesitate to raise your voice in protest against injustice. Who comes to march?

Let us go to the Mega March on July 7 to Zócalo Angel on 3:00 pm. spread the word

This paper proposes a spatiotemporal event forecasting system that leverages Twitter to promptly discover new events occurring. Methods utilize tweets as real-time and ubiquitous social sensors. Social events can be categorized into two main types: common events and unique events. Current researches into the analysis of Twitter-based streams have focused on different types of events, such as elections, disease outbreaks, and crime. The proposed generative model and associated parameter estimation 2 reviews existing work. Section 3 describes the methods by 38% and 67% on two different datasets. Extensive experimental performance evaluation demonstrates that the model inference is formalized as the observed tweet volume to the occurrence of future events. By modeling geographical priors effectively, our approach to discover the most frequently mentioned play-terms and disease outbreaks use a classifier to extract topic-centric topics by utilizing methods such as support vector machines and logistic regression; 3) Time series-based methods. This thread considers the temporal thread maps simple predictive features such as sentiment score or tweet volume to the occurrence of future events.

Figure from Zhao, Chen, Lu, Ramakrishnan, SDM 2015
Three Stories

- Face detection and recognition
- Recommendation
- News event prediction

- But not every use of machine learning is successful…
Wall Street in the mid-1980s turned to the quants – brainy financial engineers – to invent new ways to boost profits. They and their managers, though laziness and greed, built a huge financial bubble on foundations that they did not understand. It was a recipe for disaster. The journalist Felix Salmon won the American Statistical Association’s Excellence in Statistical Reporting Award for 2010. We reprint his article, first published as the cover story of Wired magazine, because it brilliantly conveys complex statistical concepts.
In the years before 2008, it was hardly unthinkable that a math wizard like David X. Li might someday earn a Nobel Prize. After all, financial economists – even Wall Street quants – have received the Nobel in economics before, and Li’s work on measuring risk has had more impact, more quickly, than previous Nobel Prize-winning contributions to the field. Yet, as dazed bankers, politicians, regulators, and investors survey the wreckage of the biggest financial meltdown since the Great Depression, Li is probably thankful he still has a job in finance at all. Not that his achievement should be dismissed. He took a notoriously tough nut – determining correlation, or how seemingly disparate events are related – and cracked it.

A formula in statistics, misunderstood and misused, has devastated the global economy.
\[ \Pr[T_A < 1, T_B < 1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma) \]

The formula that killed so many pension plans: David X. Li's Gaussian copula, as first published in 2000. Investors exploited it as a quick – and fatally flawed – way to assess risk.

### Probability

Specifically, this is a joint default probability – the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

### Survival times

The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

### Equality

A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.

### Copula

This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

### Distribution functions

The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

### Gamma

The all-powerful correlation parameter, which reduces correlation to a single constant – something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.
Machine Learning Research
Research Questions

• How do we know if the models we learn are good? Or fair?

• How do we learn from lots of data **efficiently**?

• How do we learn about more complex ideas like natural language, social networks, or biology?
@janedoe1: thanks, Obama

@johnsmith3: haha @janedoe1
Many applications have complex structure that traditional machine learning tools can’t handle. We need new algorithms and new theory.
0.2: Major(A, M) ∧ Roommates(A, B) → Major(B, M)
2.1: Major(A, M) ∧ Friends(A, B) → Major(B, M)
1.4: Studies(A, C) ∧ Dept(C, D) → Major(A, D)
Course Goals

• Primer on the science and research of machine learning

• Understand the mathematics and theory behind methods

• Flipped format
  • Watch lectures online

• Solve problems, discussion, work on homework in class sessions