Machine learning seems interesting.

I am gonna have a look at that.

Maths

never again
Learning Goals

Understand the AI-enhanced software development processes in practices
Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

From the October 2010 Issue

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren’t seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don’t know anyone, so you just stand in the corner sipping your drink—and you probably leave early.”

Goldman, a PhD in physics from Stanford, was intrigued by the linking he did see going on and by the richness of the user profiles, it all made for messy data and unwieldy analysis, but as he began exploring people’s connections, he started to see possibilities. He began forming theories, testing hunches, and finding patterns that allowed him to predict whose networks a given profile would land in. He could imagine that new features capitalizing on the heuristics he was developing might
Data-driven decision making has increased the demand for data scientists with statistical knowledge and skills.
What Do Data Scientists Work on?

**Performance Regression**
Are we getting better in terms of crashes or worse? [P3]

**Requirements Identification**
If you see the repetitive pattern where people don’t recognize, the feature is there. [P3]

**Root Cause Analysis**
What areas of the product are failing and why? [P3]

**Bug Prioritization**
Oh, cool. Now we know which bugs we should fix first. Then how can we reproduce this error? [P5]

**Server Anomaly Detection**
Is this application log abnormal w.r.t. the rest of the data? [P12]

**Failure Rate Estimation**
Is the beta ready to ship? [P8]

**Customer Understanding**
How long do our users use the app? [P1]
What are the most popular features? [P4]

**Cost Benefit Analysis**
How many customer service calls can we prevent if we detect this type of anomaly? [P9]
## Activities of Data Scientists

<table>
<thead>
<tr>
<th>Collecting</th>
<th>Building the data collection platform</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Injecting telemetry</td>
</tr>
<tr>
<td></td>
<td>Building the experimentation platform</td>
</tr>
<tr>
<td>Analyzing</td>
<td>Data merging and cleaning</td>
</tr>
<tr>
<td></td>
<td>Sampling</td>
</tr>
<tr>
<td></td>
<td>Shaping, feature selection</td>
</tr>
<tr>
<td></td>
<td>Defining sensible metrics</td>
</tr>
<tr>
<td></td>
<td>Building predictive models</td>
</tr>
<tr>
<td></td>
<td>Defining ground truth</td>
</tr>
<tr>
<td></td>
<td>Hypothesis testing</td>
</tr>
<tr>
<td>Using and</td>
<td>Operationalizing models</td>
</tr>
<tr>
<td>Disseminating</td>
<td>Defining actions and triggers</td>
</tr>
<tr>
<td></td>
<td>Applying insights/models to business</td>
</tr>
</tbody>
</table>
Data Scientist Working Styles

• Insight Providers
• Modeling Specialists
• Platform Builders
• Polymaths
• Team Leaders
Please list up to five questions you would like [a team of data scientists who specialize in studying how software is developed] to answer.
What metrics are the best predictors of failures?

If I increase test coverage, will that actually increase software quality?

What is the data quality level used in empirical studies and how much does it actually matter?

Are there any metrics that are indicators of failures in both Open Source and Commercial domains?

I just submitted a bug report. Will it be fixed?

Should I be writing unit tests in my software project?

How can I tell if a piece of software will have vulnerabilities?

Is strong code ownership good or bad for software quality?

Do cross-cutting concerns cause defects?

Does Distributed/Global software development affect quality?

Does Test Driven Development (TDD) produce better code in shorter time?
How would you approach these questions with data?

• Where to focus testing effort?
• Is our review practice effective?
• Is the expensive static analysis tool paying off?
• Should we invest in security training?
• What is a good team size?
Requirement Engineering
Engineering breakthrough: IBM introduces Watson AI for RQA

By Maggie Mae Armstrong | 2 minute read | February 28, 2019

Watson AI uses natural language processing and understanding to analyze a requirement’s text, suggesting improvements that leverage industry best practices for writing high quality requirements, based on the INCOSE Guidelines for Writing Good Requirements.

https://www.youtube.com/watch?v=pXaKgAn7PJo
Requirements Management ALM platform

Visure Requirements is an easy and comprehensive Requirements Management tool. It integrates in
the entire ALM process, providing a complete solution.

Meet Alice: Your Cognitive Assistant for Business Analysis

Your requirements gathering is about to get easier, better and faster. How? Artificial intelligence (AI) and machine
learning. The most tedious part of the requirements process can often be gathering and elicitation. Yet that part
of the process is well-suited for AI's capabilities.
Requirement Analysis

• Detection of Hidden Feature Requests from Massive Chat Messages via Deep Siamese Network. Shi et al. (ICSE), 2020
Design
Meet AIDA: Your Artificial Intelligence Design Assistant

One of the big topics in design right now is artificial intelligence. Can a computer program actually design a website? Can it help a person speed up or improve the process?

Bookmark is taking the theory to a whole new level with its Artificial Intelligence Design Assistant, or AIDA for short. AIDA learns your needs and desires and uses this knowledge to create the perfect website for you. Today we’re taking a look at how it works!

The Ultimate Designer Toolkit: 2 Million+ Assets
Coding
Code Completion

Give your development team AI superpowers

Codota automatically learns the patterns and rules in your company’s proprietary code and makes sure your developers have the best code insights, whenever and wherever they need them.

Kite adds AI powered code completions to your code editor, giving developers superpowers.

```
import os
import sys

def count_py_files_in_repos(dirname):
    |
```
https://www.youtube.com/watch?v=bF50YPyUKTQ
Use machine-learning-assisted code completion

You can utilize machine learning models to rank most suitable items higher in the suggestions list.

To do this, in the Settings/Preferences dialog, go to Editor | General | Code Completion and enable the Rank completion suggestions based on Machine Learning option under Machine Learning-Assisted Completion.

This feature is experimental, so ranking may not change noticeably.
loss = tf.reduce_sum(tf.square(linear_model - y))
optimizer = tf.train.GradientDescentOptimizer(0.01)

train = optimizer
Deep TabNine: A Powerful Code Autocompleter For Developers

https://medium.com/syncedreview/deep-tabnine-a-powerful-ai-code-autocompleter-for-developers-70454a5953fe
Sharif Shameem @sharifshameem · Jul 13

This is mind blowing.

With GPT-3, I built a layout generator where you just describe any layout you want, and it generates the JSX code for you.

WHAT

Describe a layout.
Just describe any layout you want, and it'll try to render below!

```html
<h1 style={{fontSize: 50, color: 'white'}}>WELCOME TO MY NEWSLETTER</h1>
<button style={{color: 'white', backgroundColor: 'blue'}}>Subscribe</button>
```

WELCOME TO MY NEWSLETTER

Subscribe

https://twitter.com/sharifshameem/status/1282676454690451457?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1282676454690451457%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fanalyticsindiamag.com%2Fopen-ai-gpt-3-code-generator-app-building%2F
Aroma: Using machine learning for code recommendation
Introducing GitHub Copilot: your AI pair programmer
Debugging

Relative cost to fix bugs, based on time of detection

https://deepsources.io/blog/exponential-cost-of-fixing-bugs/
Ubisoft: ML catches 70% of bugs prior to testing

“The statistical nature of machine learning involves us changing the way we work,” he says. Unlike traditional software, in which developers write out rules for the application to follow, machine-learning algorithms use data to guide how the software should act.

-- Yves Jacquier, executive director, production studio services, Ubisoft Montreal

https://www.pmi.org/learning/library/ai-debug-code-11523
Debugging

Getafix: How Facebook tools learn to fix bugs automatically

Quality Assurance
DeepCode

**STEP 1: PARSING**
This is the only language-specific part of our platform and enables us to add support for any custom language in a matter of weeks.

**STEP 2: SOLVERS**
(DECLARATIVE STATIC ANALYSIS)
Our custom language-independent linear-complexity Datalog solvers allow us to analyze huge repositories in a matter of seconds.

**STEP 3: ML ALGORITHMS**
Our custom Semantic Facts representations allows us to run powerful ML algorithms to understand the structure, function, and intent of the code.
Podcast:
https://www.youtube.com/watch?v=bRUpRxFFlyQ
https://www.youtube.com/watch?v=pYzfGaLTqC0

CodeQL helps you explore code quickly to find and eradicate all variants of vulnerabilities before they become a problem.

By automating variant analysis, CodeQL enables product security teams to find zero-days and variants of critical vulnerabilities.
Deployment
Continuous app delivery firm Harness raises $60M

BY MIKE WHEATLEY

https://siliconangle.com/2019/04/23/continuous-app-delivery-firm-harness-raises-60m/
Project Management
How CraneAi uses Artificial Intelligence to help teams build apps faster

Behind the scenes look at how CraneAi’s uses artificial intelligence to empower teams

Ryan Hickman  Follow
Nov 30, 2018 · 5 min read
“Tara’s mission is to help teams develop their plans with visibility and predictability.”

COMPANY NEWS
Ford and Cisco are turning to an AI company to find the best freelance programmers

EDITOR
January 18, 2018
Trade-off?
I sometimes see people refer to neural networks as just “another tool in your machine learning toolbox”. They have some pros and cons, they work here or there, and sometimes you can use them to win Kaggle competitions. Unfortunately, this interpretation completely misses the forest for the trees. Neural networks are not just another classifier, they represent the beginning of a fundamental shift in how we write software. They are Software 2.0.

https://medium.com/@karpathy/software-2-0-a64152b37c35
I got myself a cool AI T-Shirt - then the sticker began to peel off
Data Analytics for Software Engineering (DA4SE)

Software Refactoring
- Refactoring Field Study
- Quantifying Refactoring Cost and Benefits
- Impact on Regression Testing
- Role of API Refactoring

API Evolution
- Role of API Refactoring
- API Stability

API Usage Mining
- ExampleCheck: API Misuse Detection
- ExampleExplore: Visualizing API Examples at Scale
- ExampleStack: API Usage Adaptation

Program Transformation from Examples
- Sydit
- LAS
- Cookbook
- Bug Finding
- Refactoring Bugs
- Cloning: Inconsistencies
- Fault Tracer
- Modularity Violations
- Prioritizing Tests for Refactoring

Empirical Studies of Software Changes
- Code Redundancy
  - Clone genealogy
  - Copy and paste practices
  - Long lived clones
  - Software forking and code porting

- Software Patches
  - Supplementary patches
  - Omission errors

- Logical Differentiation
  - Lsdiff
  - Vdiff for VHDL

- Refactoring Reconstruction
  - RefFinder
- API Usage Adaptation
  - LbSync
  - AURA
  - API Matching

- Interactive Code Review
  - Critics

- Test Transplantation
  - Grafter

- Clone Removal Refactoring
  - RASE

Recommendation Systems

Automated and Interactive Software Dev Tools