

What is the big in (Big) data? And how lean can your data be...

#AIMonday - Berlin, 17.09.2018

Dat Tran - Head of Data Science @datitrn

echo \$(whoami)

idealo

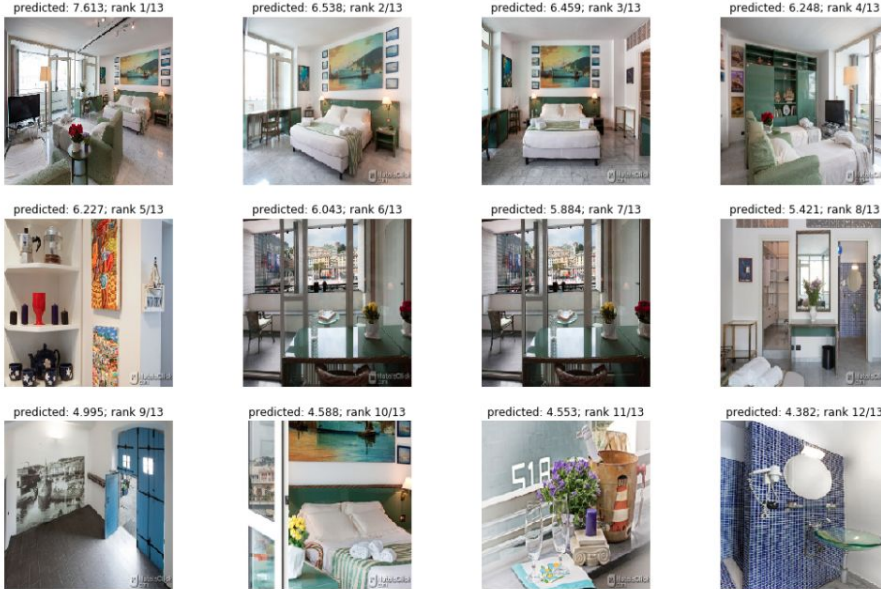
idealo

 Pivotal **Labs**

accenture[>]digital



What do we do at idealo? Some examples...



Hotel image ranking for both aesthetic and technical quality



Low-to-high resolution

Das könnte Dich auch interessieren



Recommendation engine

Check us out!



<https://github.com/idealo>



<https://medium.com/idealo-tech-blog>

Let me start with a story...



This is the reality...



The story goes on...



The story ends like this...



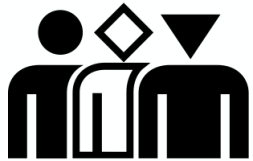
Moral of the story



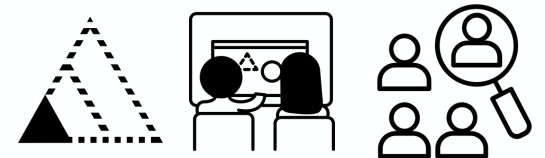
- A lot of data initiatives fail because there are people on power who have no clue what they do



- Spending a lot of money and hiring expensive consultancy don't really help

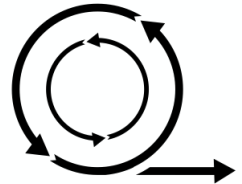


- Not able to create a sustainable risk taking and **agile culture**



What is MVP?

Eric Ries says Minimum Viable Product (MVP) is essentially a “**version of a new product** which allows a team to collect the **maximum amount of validated learning about customers** with the **least effort**”.



Why is MVP important?

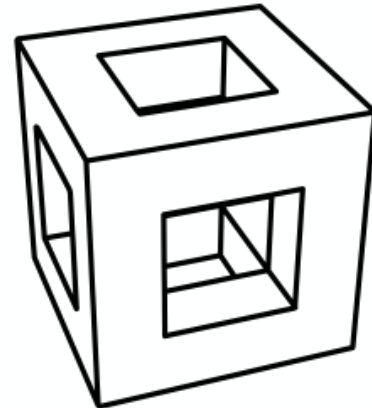


How does MVP relate to ML products?

→ Machine learning ends up as **data products**, so they can be also considered like **(normal) products**

Three Dimensions:

1. Minimum Viable Model
2. Minimum Viable Platform
3. Minimum Viable (Data) Product



MVP For Self-Driving Cars



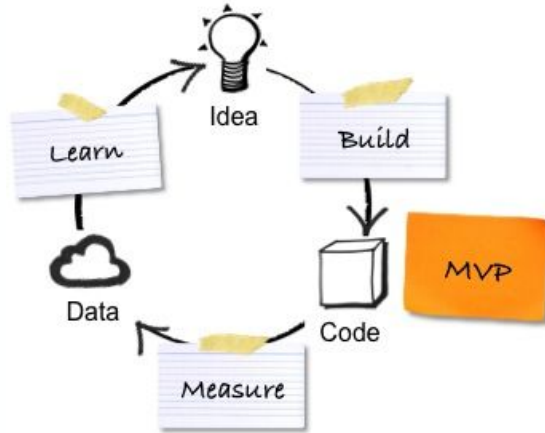
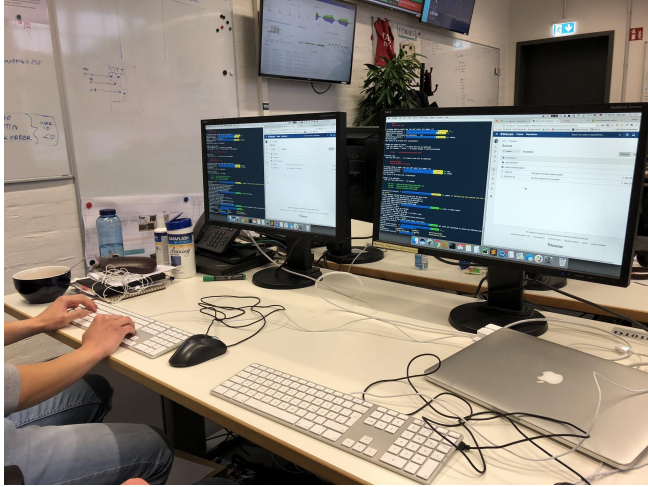
Who knows what this is...?



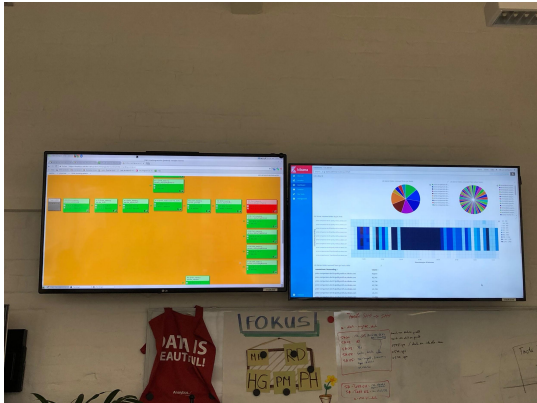
Digital Labs are Cargo Cult



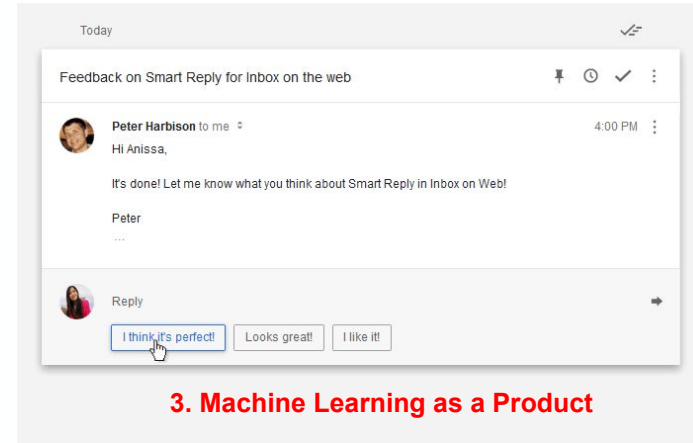
Culture is the Real Reason



2. Hypothesis-driven experimentation



1. Engineering culture: Pair Programming, CI/CD, TDD



3. Machine Learning as a Product

The Right People...



Would you apply for this job?

Senior Data Scientist Job Ads from Deutsche Telekom Bonn

IHRE AUFGABEN

- Evaluierung, Entwicklung und Verwaltung von Datenanalyseanwendungen und Kundenlösungen
- Anwendung von Machine Learning und analytischen Ansätzen für die Entwicklung von IoT-Lösungen
- Entwicklung von Algorithmen, Konzipierung von Lösungen und Kooperation mit Domäne-Experten und Data-Engineers
- Unterstützung bei Proof of Concepts in hochmodernen Datenanalytik-Umgebungen, um die Umsetzbarkeit aufzuzeigen und Lösungen für (IoT) datenbezogene Geschäftsanforderungen zu entwickeln
- Unterstützung des Digital Division (DD) Vertriebs bei Kundenworkshops/Diskussionen um geeignete Ansätze, Methoden, Werkzeuge und Lösungen zu spezifizieren
- Entwerfen und Implementieren statistischer Datenqualitätsverfahren für neue Datenquellen
- Interesse><http://www.telekom.jobs/findedeinteam/#/datanistas>

IHR PROFIL

- Dokortitel oder Master in Data Science, Physik, Mathematik, Statistik oder Informatik
- Relevante Berufserfahrung (mehr als 10 Jahre Berufserfahrung in IoT-Industrie)
- Erfahrung in (funktionalen) Programmiersprachen, z.B. Python, Java, R, Scala, Javascript
- Software Engineering Know-How in agilen Methoden für alle Schritte (Design, Architektur, Entwicklung, Test, Einsatz)
- Erste Erfahrungen in interaktive Datenvisualisierung mit Datenanalytik - Tools (z.B. Tableau, Visual Analytics)
- Kenntnisse in Mathematik und statistischen Methoden
- Professionelle Fähigkeiten mit Datenanalyse-Tools (z.B. Knime, Rapidminer, Talend, Tableau, Cassandra, Hadoop, Spark), Kompetenzen zur Gestaltung und Verwaltung von Big Data Umgebungen (z.B. Hadoop), Auswertung und Konfiguration von Komponenten (z.B. Spark, DB), Anpassung und Entwicklung von Algorithmen für Kundenlösungen
- Erfahrungen mit zwei der folgenden Anwendungsbereiche: IoT Analysen, Zeitreihenanalytik, Social Media Analysen, Standortdatenanalytik, Sicherheitsdatenanalytik, Kundendatenanalytik, Signalverarbeitung, Anomalie Erkennung



Summary

- There is no right or wrong - do what works for you - fail fast
- Establish a real iterative MVP culture for data
- Hiring is difficult but you (as a company) need to change to get the best people on board



Questions?



Url: www.dat-tran.com
Twitter: @datitran