AI and Data Science Approaches Adoption for the Innovation of Emerging Economies

Seid Muhie Yimam (Ph.D.)
HCDS, UHH, Germany

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

HOUSE OF COMPUTING & DATA SCIENCE

Outline

01

AI & Data Science

- Trends
- Challenges

02

Use-cases

- Digital humanity
- Musicology
- Legal documents

03

Digitalization for Data Science in Emerging Economy

- Digitalization
- Issues
- Way forward

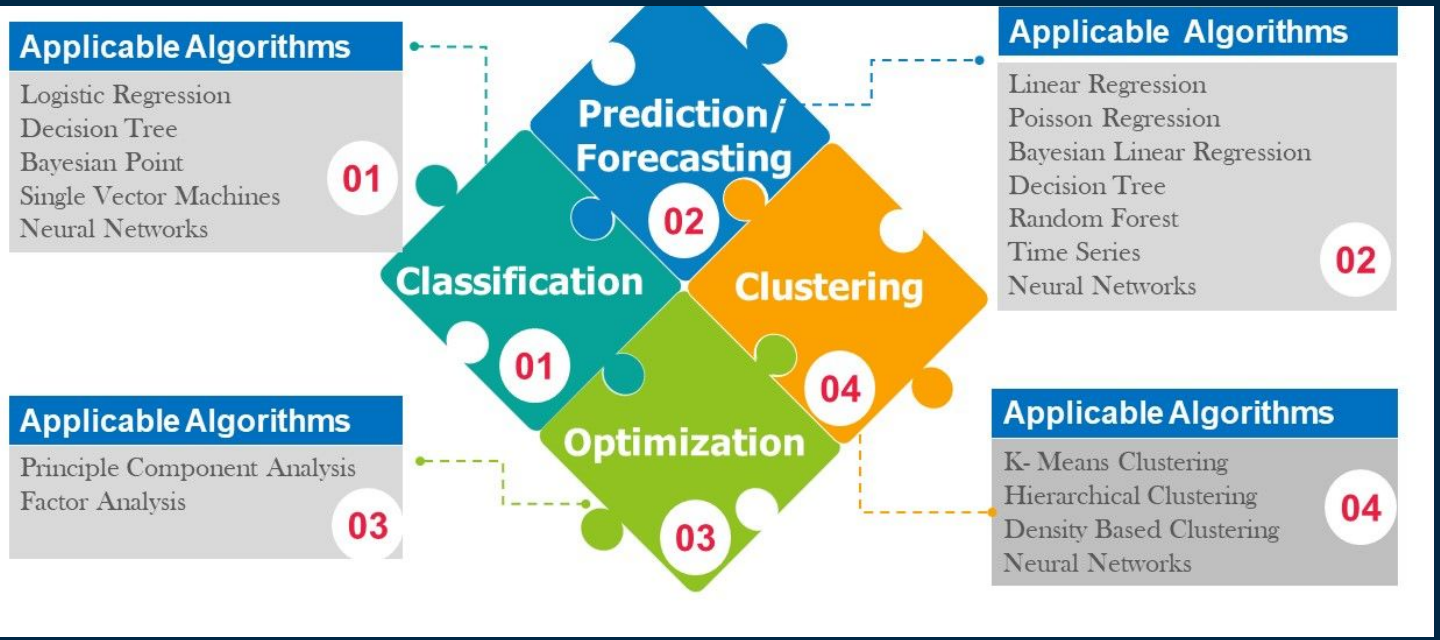
01 - AI & Data Science

AI and Data Science

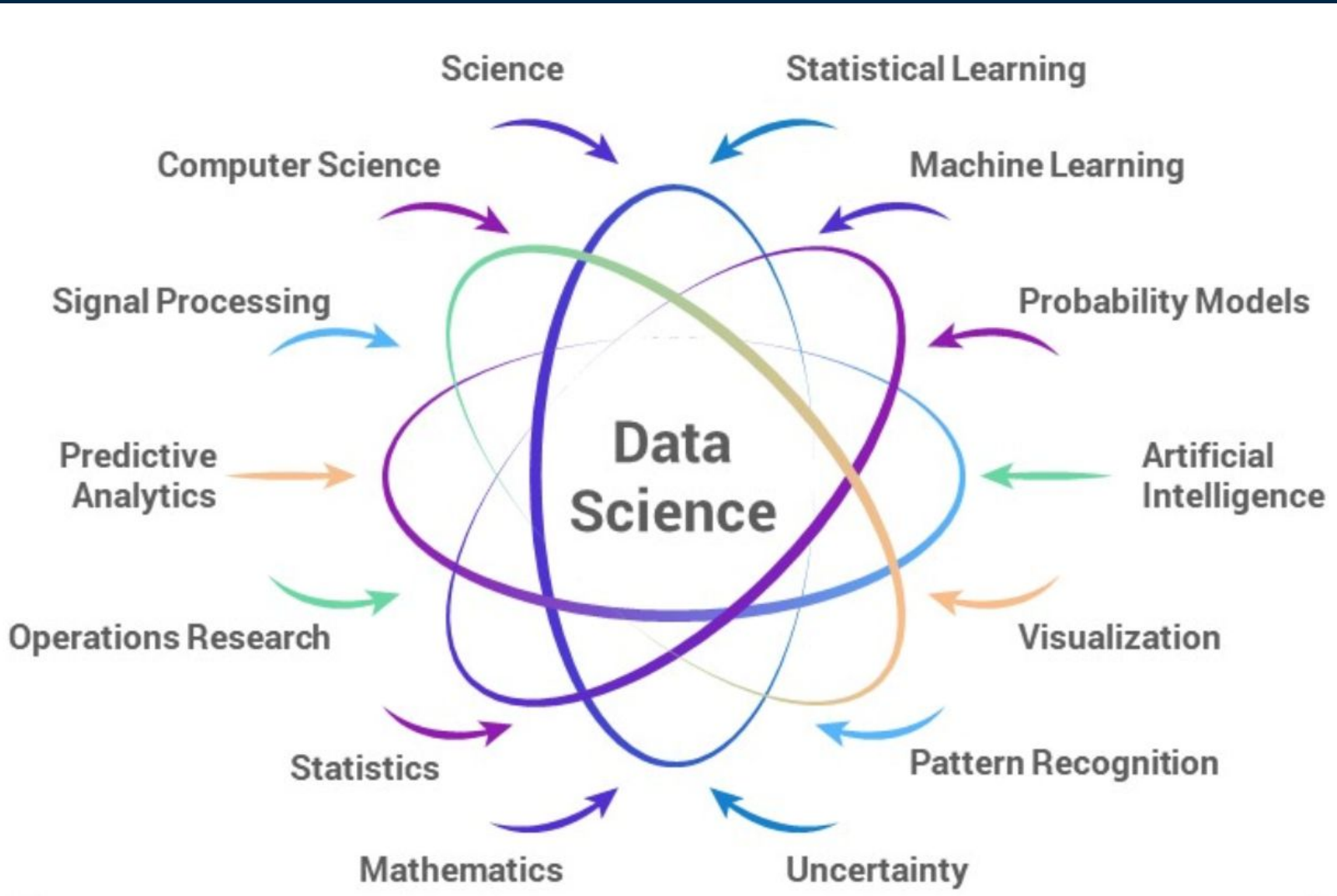
- Artificial Intelligence - Systems that display intelligent behavior by analysing their environment and taking action - with some degree of autonomy - to achieve a specific goal - Philip Boucher

- Data Science: The process of collecting, storing, segregating and analyzing data which serves as a valuable resource for organizations to carry out data-driven decision making - Edureka

Machine Learning is a subsection of Artificial intelligence that devices a way by which systems can automatically learn and improve performance from experience.

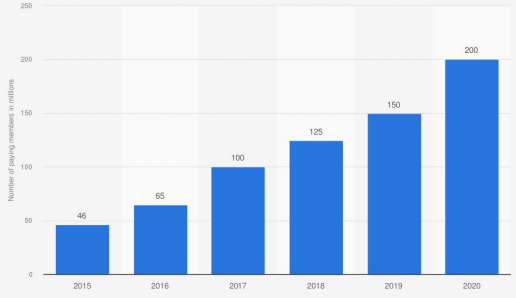


https://blogs.sap.com/wp-content/uploads/2019/07/ML-Algorithm.jpg


Science
Statistical Learning
Computer Science
Machine Learning
Signal Processing
Probability Models
Predictive Analytics
Data Science
Artificial Intelligence
Operations Research
Visualization
Statistics
Pattern Recognition
Mathematics
Uncertainty

Trends in Data ...

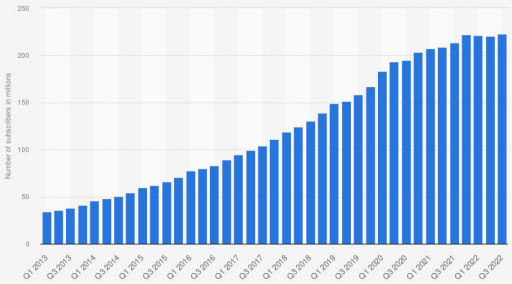
Number of paying Amazon Prime members worldwide from 2015 to 2020 (in millions)



Sources
Amazon; Marketplace Pulse
© Statista 2022

Additional Information:
Worldwide; Amazon; 2015 to 2020

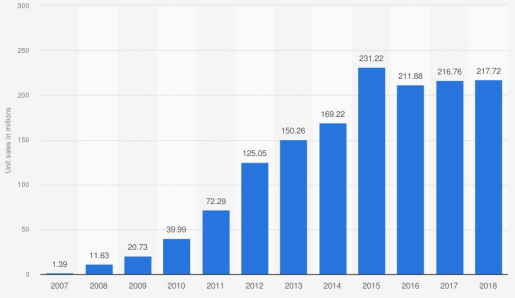
Number of Netflix paid subscribers worldwide from 1st quarter 2013 to 3rd quarter 2022 (in millions)



Source
Netflix
© Statista 2022

Additional Information:
Worldwide; Q1 2013 to Q3 2022; excluding free trials

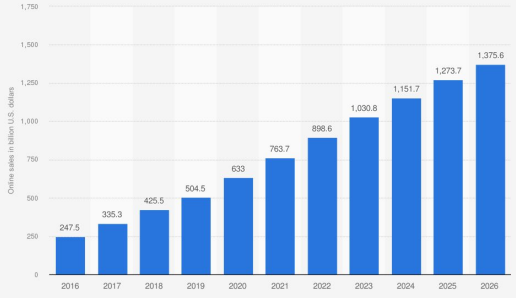
Unit sales of the Apple iPhone worldwide from 2007 to 2018 (in millions)



Source
Apple
© Statista 2022

Additional Information:
Worldwide; 2007 to 2018

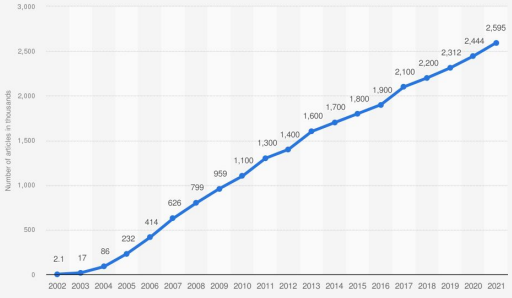
E-commerce net sales of Alibaba Group from 2016 to 2026 (in billion U.S. dollars)



Source
Edge by Ascential
© Statista 2022

Additional Information:
Worldwide; January 2022; estimates

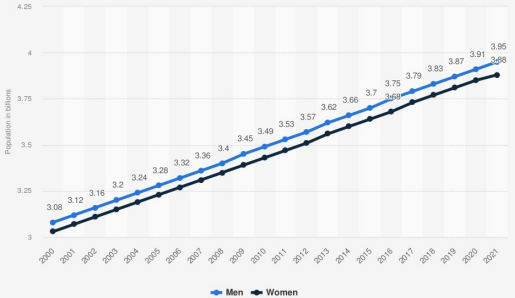
Number of German-language Wikipedia articles from 2002 to 2021 (in 1,000)



Source
Wikipedia
© Statista 2022

Additional Information:
Germany; Wikipedia; 2002 to 2021; April each year; 2019: June 6th

Global population from 2000 to 2021, by gender (in billions)



Source
World Bank
© Statista 2022

Additional Information:
Worldwide

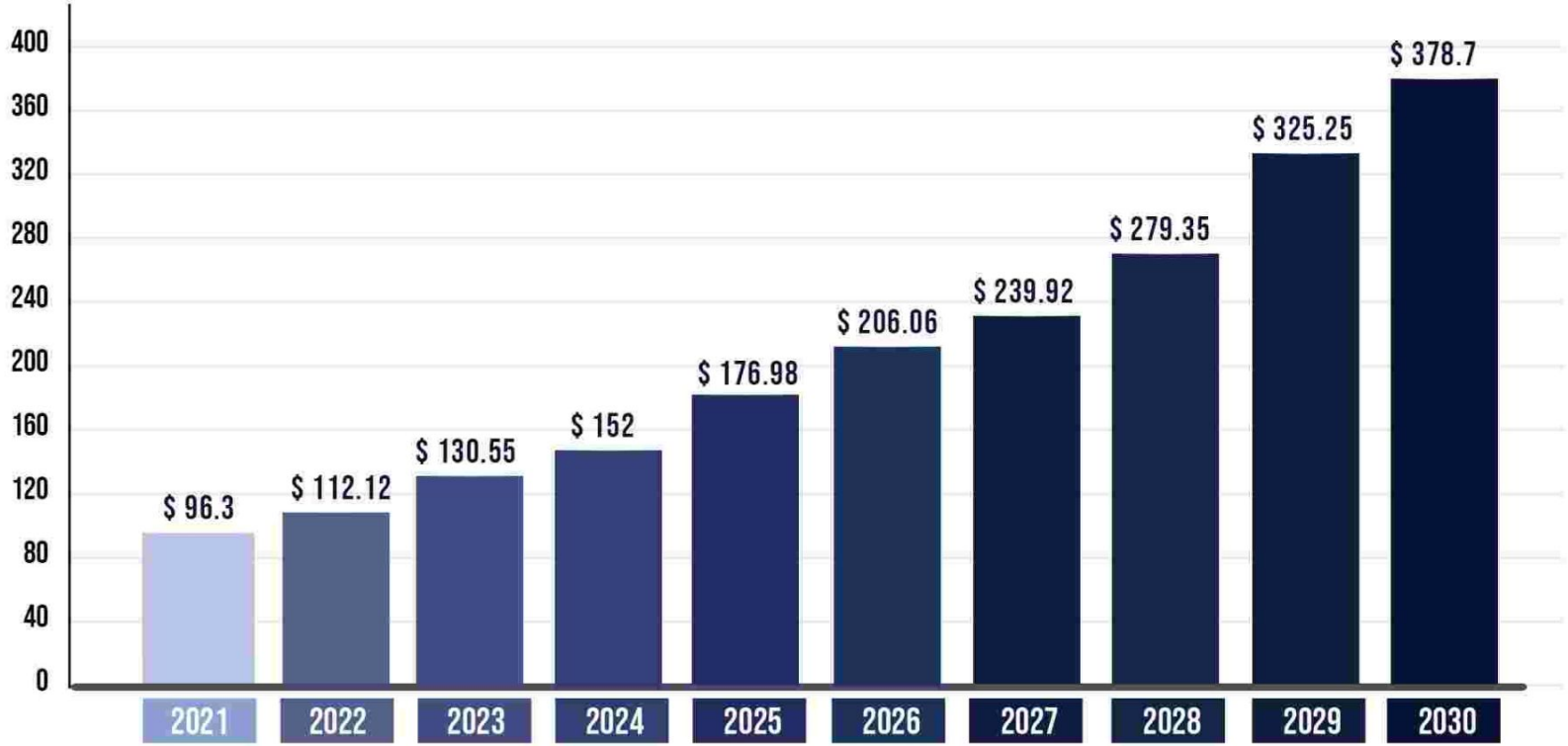
Data Science Market

Report: Global Data Science Platform Market by Component (Platform, Services), by Application (Marketing & Sales, Logistics, Finance and Accounting, Customer Support, Others), by Industry Vertical (BFSI, Retail and E-Commerce, IT and Telecom, Transportation, Healthcare, Manufacturing, Others): Global Opportunity Analysis and Industry Forecast, 2020-2030

Vikas Gaikwad , Vineet Kumar

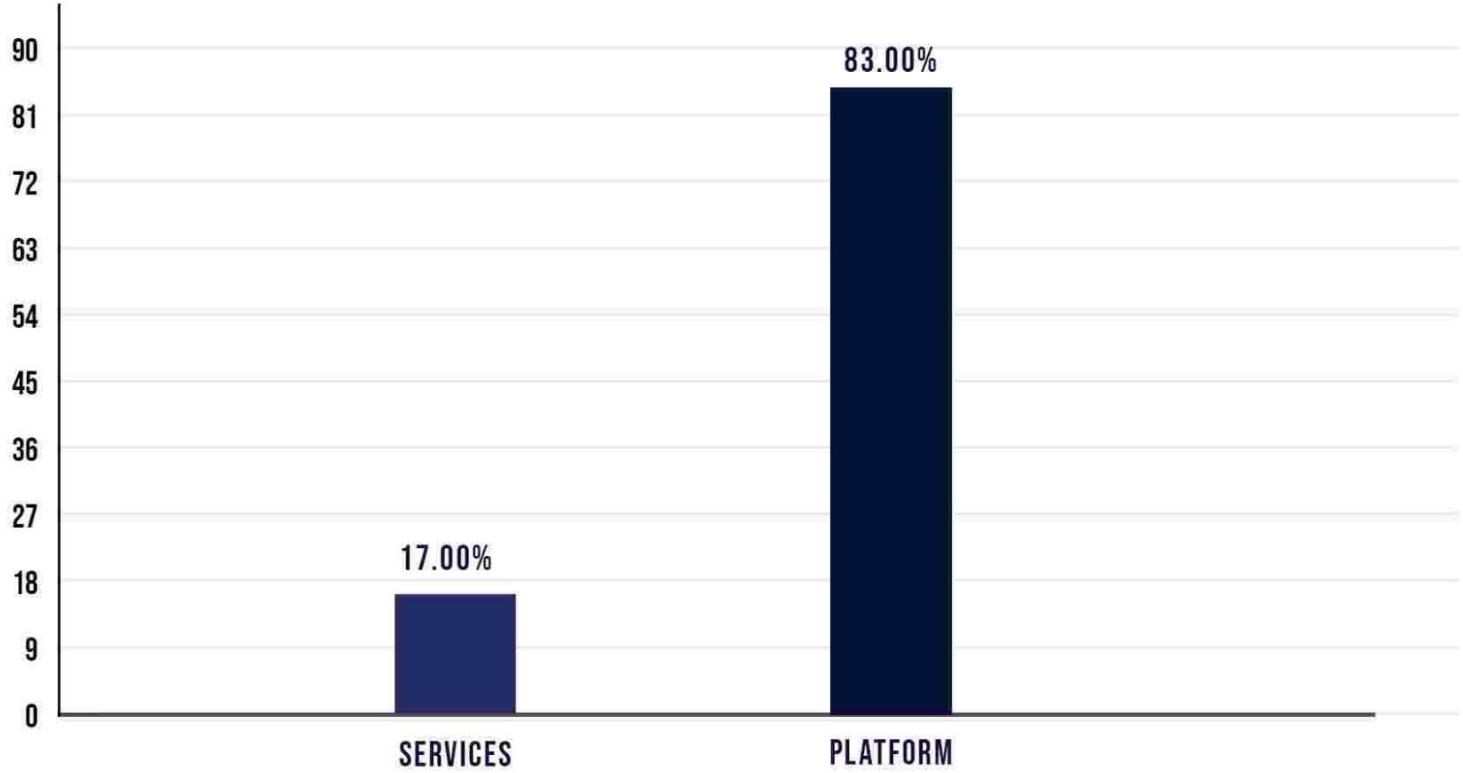
PRECEDENCE RESEARCH

DATA SCIENCE PLATFORM MARKET SIZE, 2021 TO 2030 (USD BILLION)



Source: www.precedenceresearch.com

DATA SCIENCE PLATFORM MARKET SHARE, BY COMPONENT, 2021 (%)

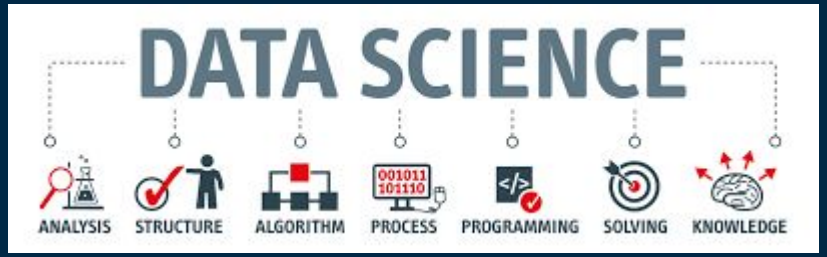


© PRECEDENCE RESEARCH

Data Science Platform Market Share, By Region, 2021 (%)

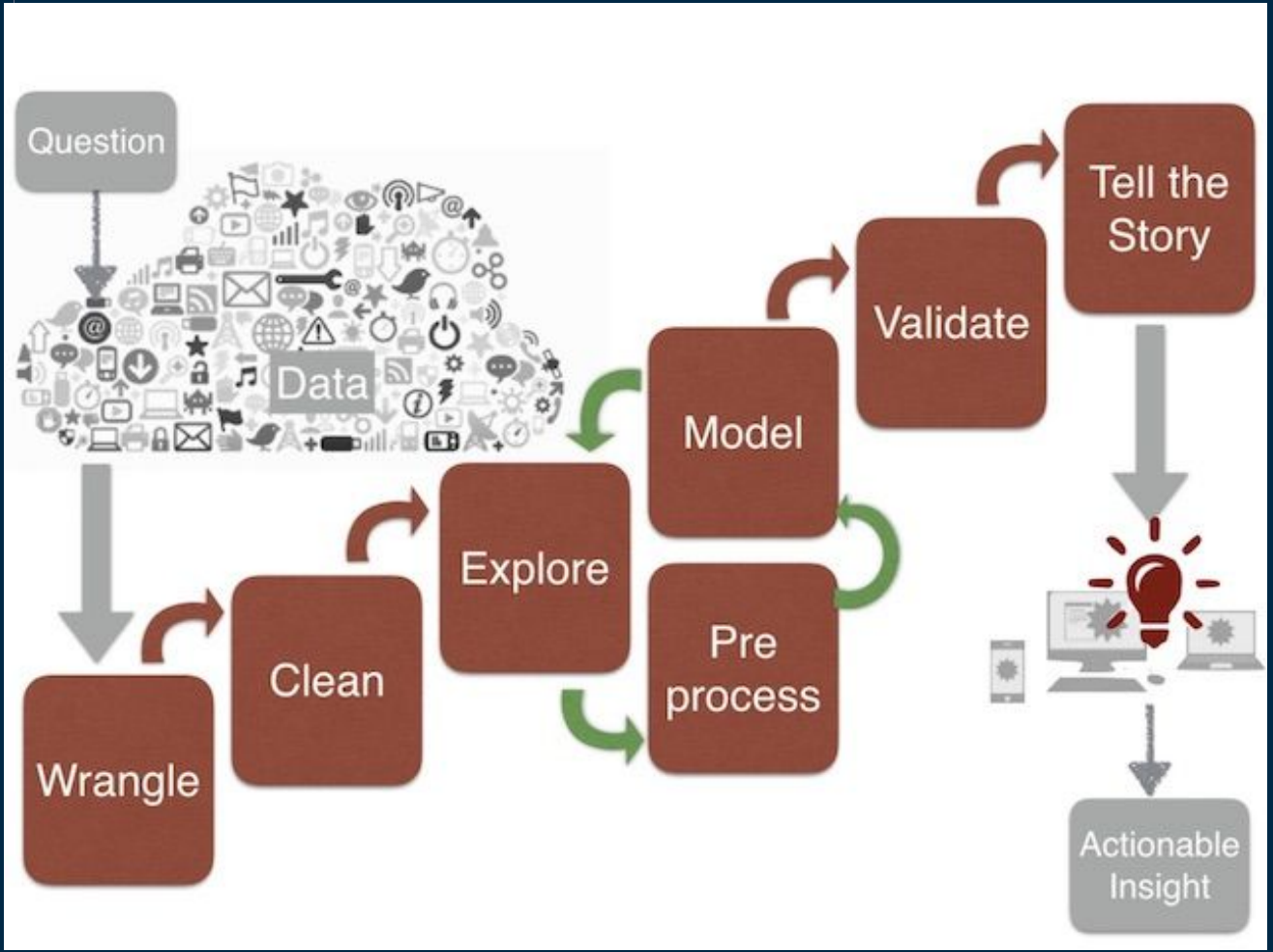
Regions	Revenue Share in 2021 (%)
North America	36%
Asia Pacific	25%
Europe	29%
Latin America	6%
MEA	4%

Data Science Pipeline



https://hevodata.com/learn/data-science-pipeline/

Data Science Pipeline

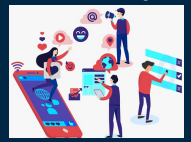


https://d8it4huxumps7.cloudfront.net/bites/wp-content/uploads/2018/09/datascience-pipeline.jpg

Data Science Enablers

Enabers

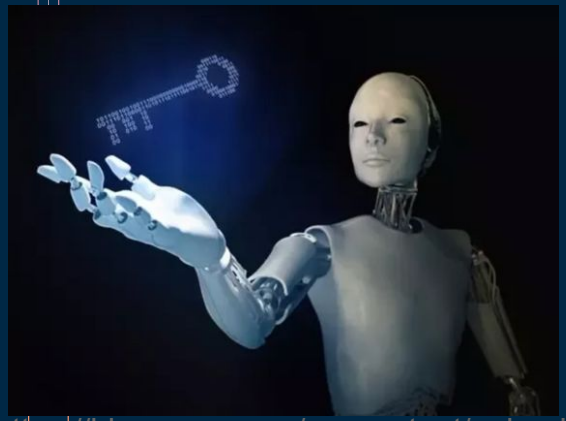
- Social, Mobile, Cloud, and IoT (SMACIT)



- Artificial intelligence, Machine learning / Deep learning

- Data Storage, computation, Visualization and Analytics

AI and Data Science Challenges



https://blogs.sap.com/wp-content/uploads/2019/07/security.png

ML and data science challenges

By Pavel Klushin

- Data collection: Large amount of unorganized data (storage is cheap), a lot of data source (difficult to find right source)
- Data security and privacy: compliance requirements
- Large data volumes: Storage, management, model building
- Missing data: lack of data documentation
- Extracting the right insights: right insight, real-time, high-quality
- Right talent: huge skills gap and talent shortage
- Expensive: right team setup, model building, infrastructure

Data Science Toolkits

Comprehensive mathematical functions, random number generators, linear algebra routines, Fourier transforms, easy to use structures and data analysis, algorithm for optimization, integration, interpolation

Machine learning, evaluation, learning algorithms…

Software environment for statistical computing and graphics, the development of both web and desktop applications

Python based data visualization, enabling data-driven decisions

02 - Use-cases at HCDS

HOUSE OF COMPUTING AND DATA SCIENCE

- A **central institution** of the University of Hamburg
- Supports **interdisciplinary research** and application of **innovative digital methods** in close cooperation with its partners from **science and industry** in the Hamburg metropolitan region.
- **Coordinates** and **supports** the implementation of the **digital strategy** in research at the University of Hamburg.

Philosophy

Generare imaginary picture for text using Deep Daze!

https://github.com/lucidrains/deep-daze

a victorian beach in a cave →

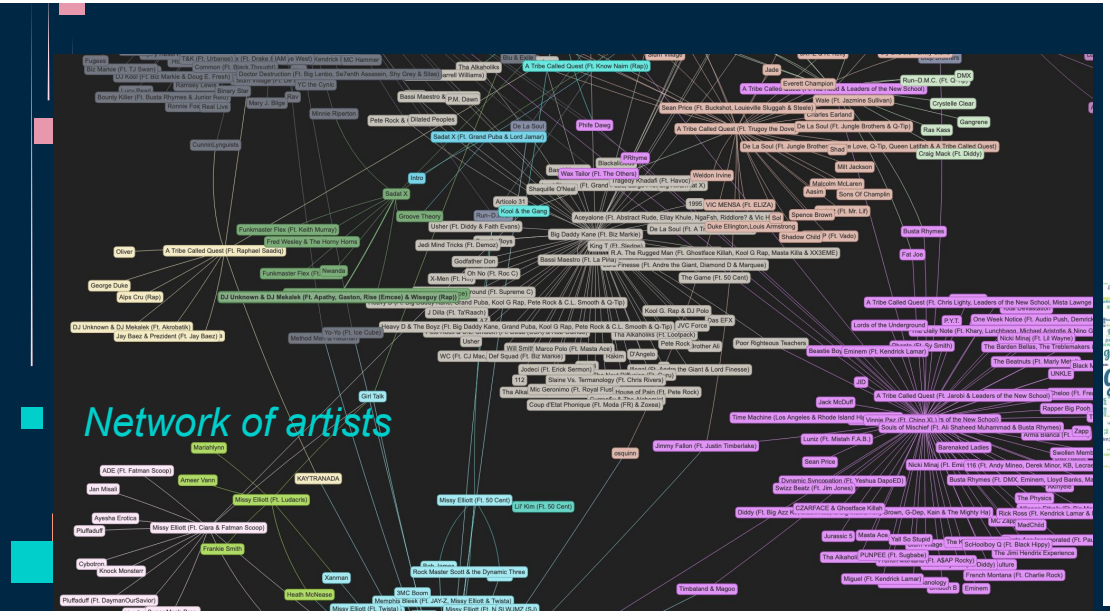
Musicology

Study different features about music of the world

- Existing APIs for music song extraction
- Extract music lyrics from Genius
- Topic modeling for songs
- Sentiment analysis for lyrics
- Network analysis of artists
- More song feature extraction from Spotify
- Lyrics Generation using GPT2

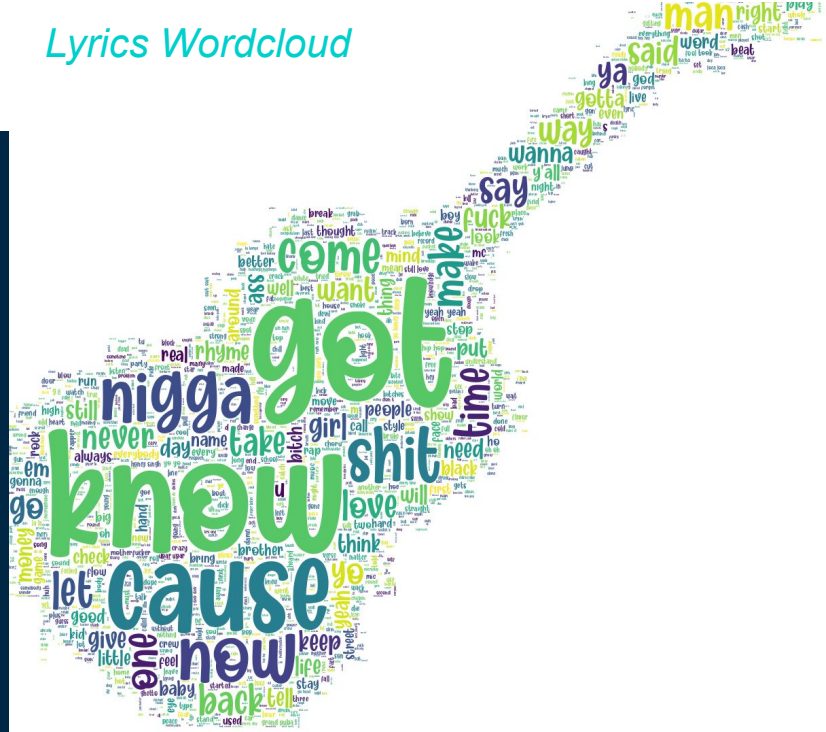
```
illuminated thoughts she dances
illuminated thoughts she dances in the pool and dances all
over the floor, giving me a good workout for my gyms.

"What do you mean you can't
```

GPT2 - Lyrics Generator



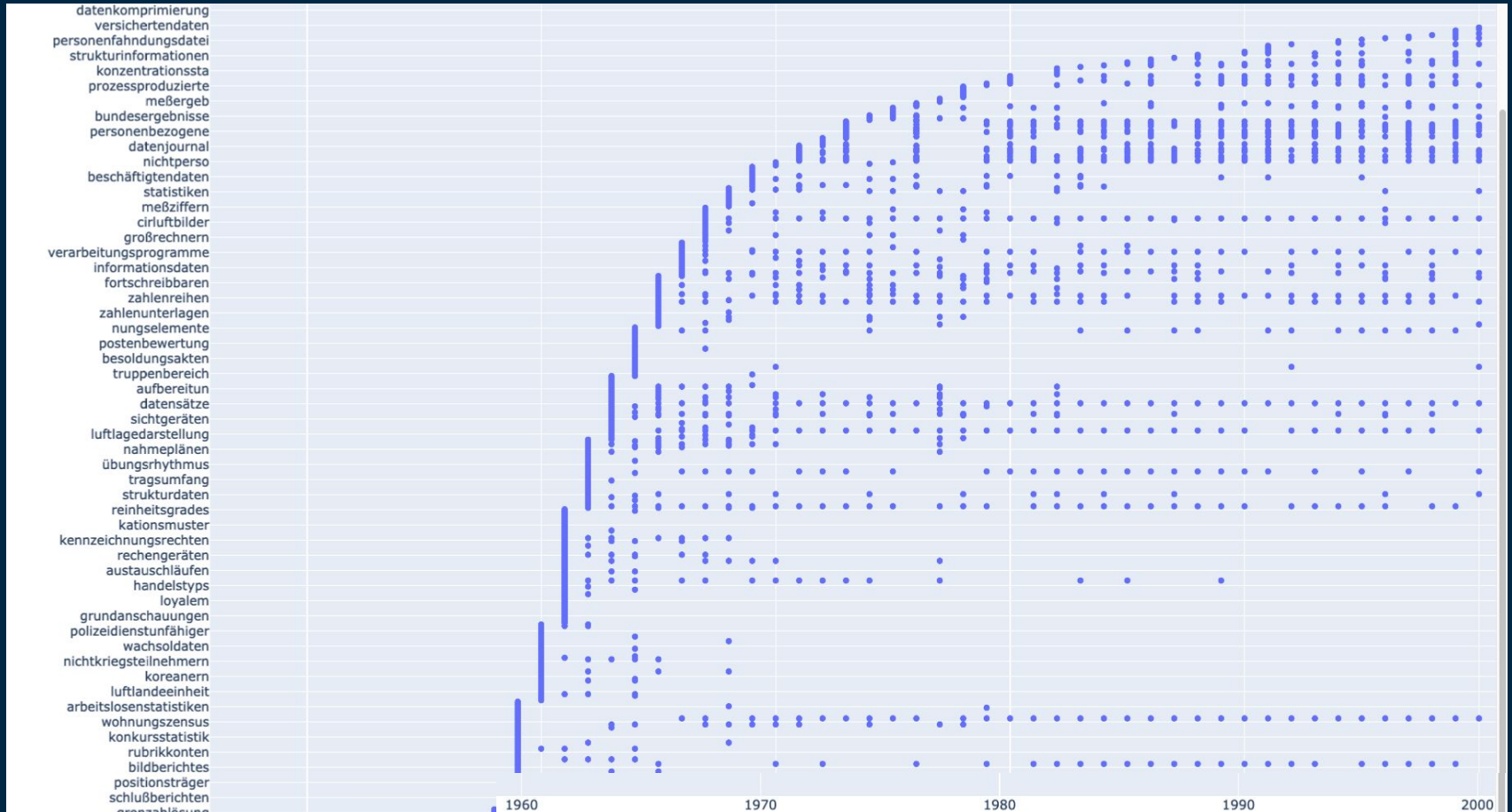
Network of artists

Lyrics Wordcloud



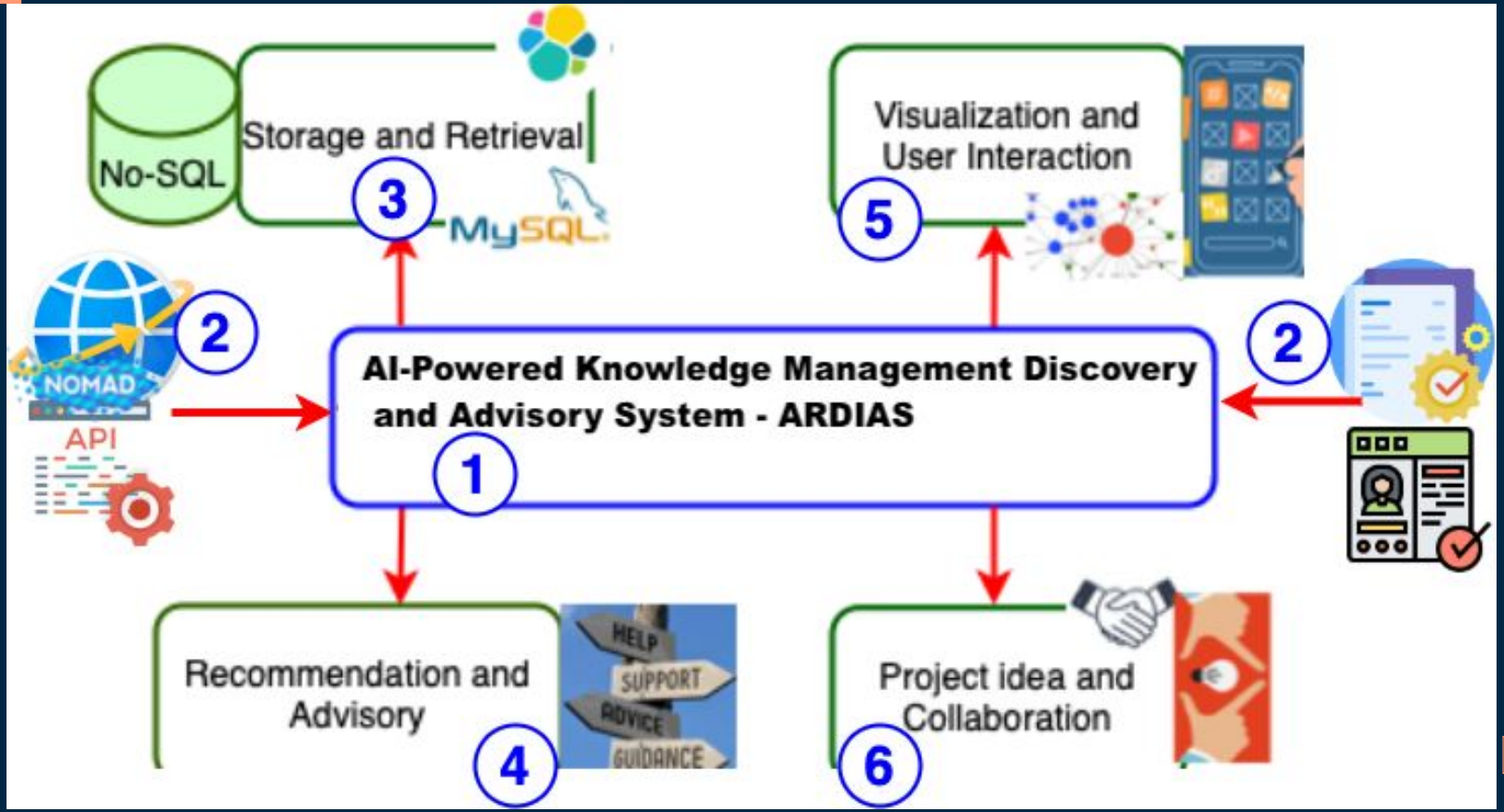
Legal Document Processing

Understand German regulations related to “Data” over time, 1950 -



ARDIAS: AI-Enhanced Research Management, Discovery and Advisory System

Improve the Findability, Accessibility, Interoperability, and Reuse of research data



Publications | Co-author Network | Research Focus

Search for works by Wei Shen here ... Sort by

Showing 10 results of 186. Fetched in 55ms.

Carbon quantum dots and their applications

Shi Ying Lim, Wei Shen, Zhiqiang Gao , **2015** - **Chemical Society Reviews**

This review covers the progress in research and development of carbon quantum dots their applications chemical sensing, biosensing, bioimaging, nanomedicine, photocatalysis electrocatalysis.

Read more

99 3097

SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation

Wei Shen, Shuai Le, Yan Li, Fuquan Hu , **2016** - **PLOS ONE**

FASTA and FASTQ are basic ubiquitous formats for storing nucleotide protein sequences. Common manipulations of FASTA/Q file include converting, searching, filtering, deduplication, splitting, shuffling, sampling. Existing tools only implement some these manipulations, not particularly efficiently, available certain operating systems. Furthermore, the complicated installation process…

Read more

Open Access 99 722

Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions

Wei Shen, Jianyong Wang, Jiawei Han , **2015** - **IEEE Transactions on Knowledge and Data Engineering**

The large number of potential applications from bridging web data with knowledge bases have led to an increase in the entity linking research. Entity is task link mentions text their corresponding entities a base. Potential include information extraction, retrieval, and base population. However, this challenging due name variations ambiguity. In survey, we present thorough overview analysis main…

Read more

99 485

Wei Shen

Jiangsu University of Science and Technology , CN

186
99 9704
Openalex: A2602767326
Orcid: 0000-0001-8159-5728
Mag: 2602767326
Scopus: authorDetails.url?authorID=57001188100&partnerID=MN8TOARS

Upcoming Workshop – Thought in humans and machines

Speakers:

Monday, November 14th & Tuesday, November 15th, 2022

- Anjali **Bhat**: “Dishbrain and immunoception: Alterity at a biological interface “ (Wellcome Centre for Human Neuroimaging, London)
- Cameron **Buckner**: Sentimental machines – artificial moral and social cognition in the age of deep learning (Philosophy, University of Houston (Zoom))
- Anne **Lauscher**: Fairness & Inclusion in modern AI: Cases for Historical Fiction? (Data Science, Universität Hamburg)
- Sara **Mann**: "Understanding via Exemplification in XAI" (Philosophy, Technical University of Dortmund)
- Dimitri Coelho **Mollo**: "Models of (Artificial) Intelligence: idealisation and behaviour" (Philosophy, Umeå University)
- Stephanie **Nelli**: The geometries of flexible knowledge structures in minds and machines (Cognitive Science, Occidental College, Los Angeles)
- Nicholas **Shea**: How to Solve and Avoid the Frame Problem (Philosophy, School of Advanced Study, London (Zoom))

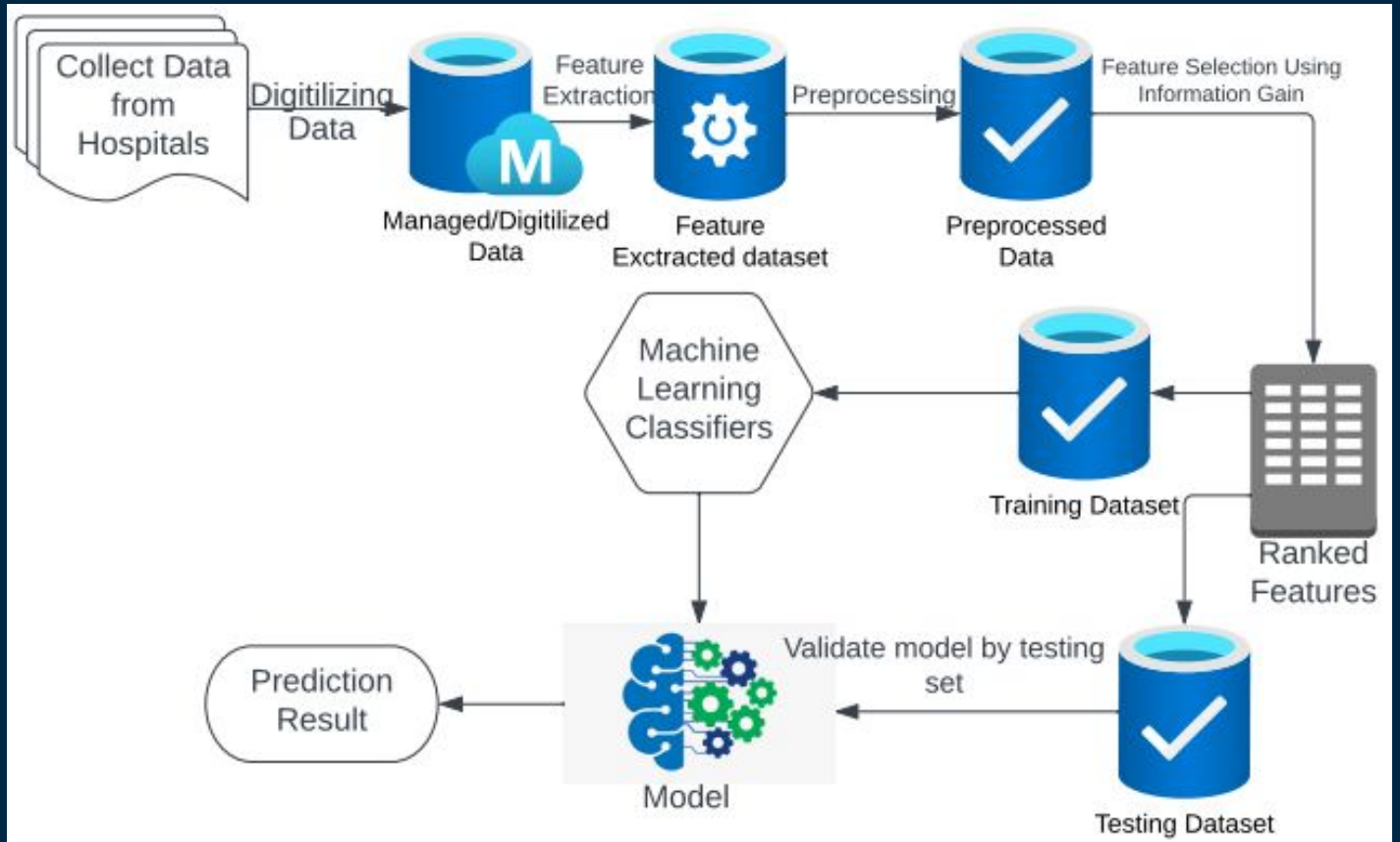
03 - Digitalization for Data Science in Emerging Economy

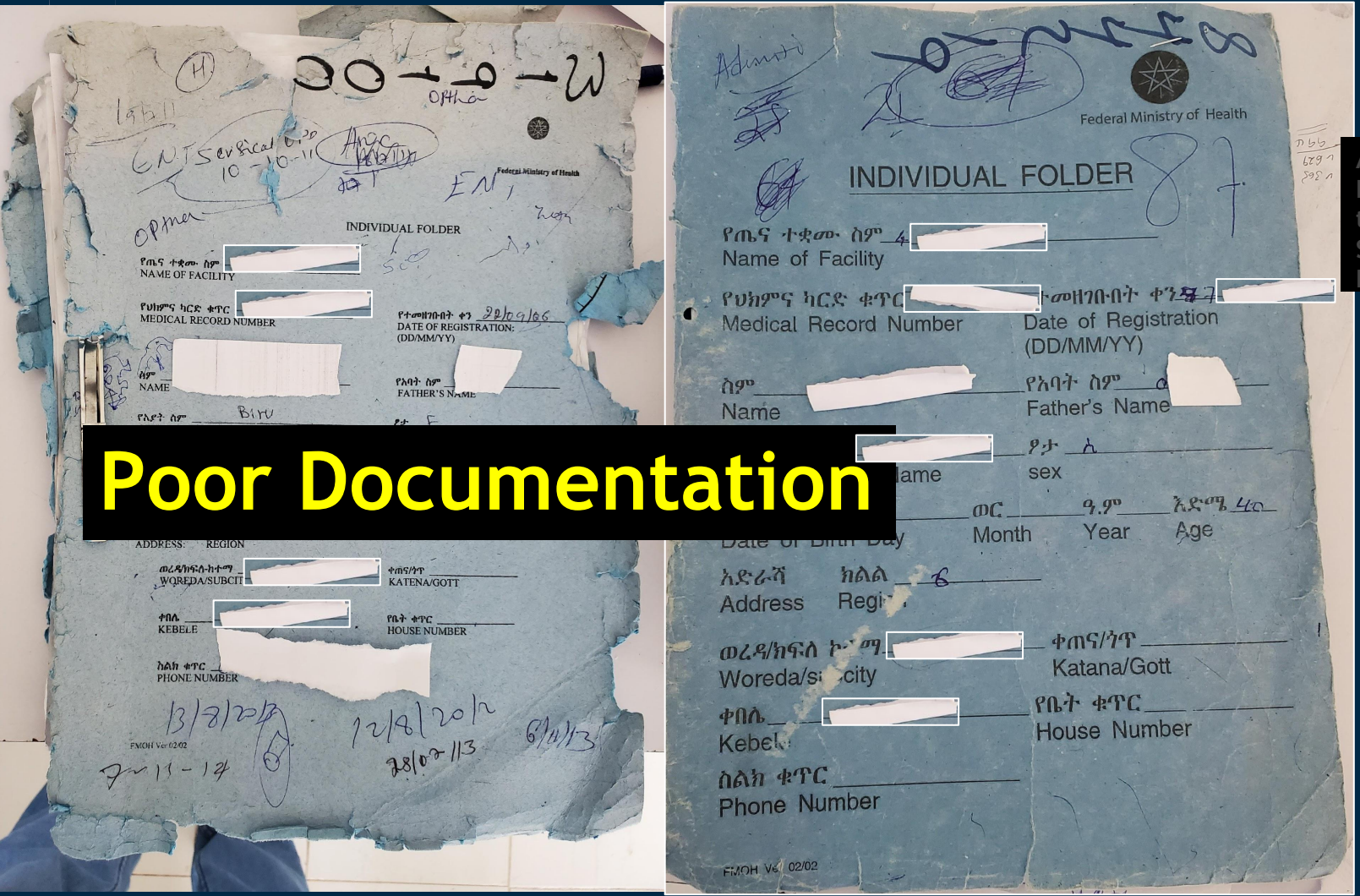
Datafication

The process that makes a business data driven is by **collecting huge data from various sources** and storing them in centralized places to find new insights that lead to better opportunities – can be termed as Datafication. Datafication will take big data analysis to new heights – into real insights, future predictions and intelligent decisions.

projectpro.io

Digitalization for Health



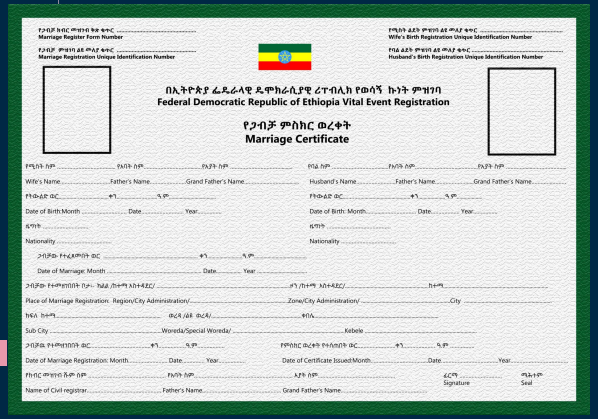


Poor Documentation

Adapted from Kerebih's Seminar talk, Applied Data Science course, BDU, BiT, 2022

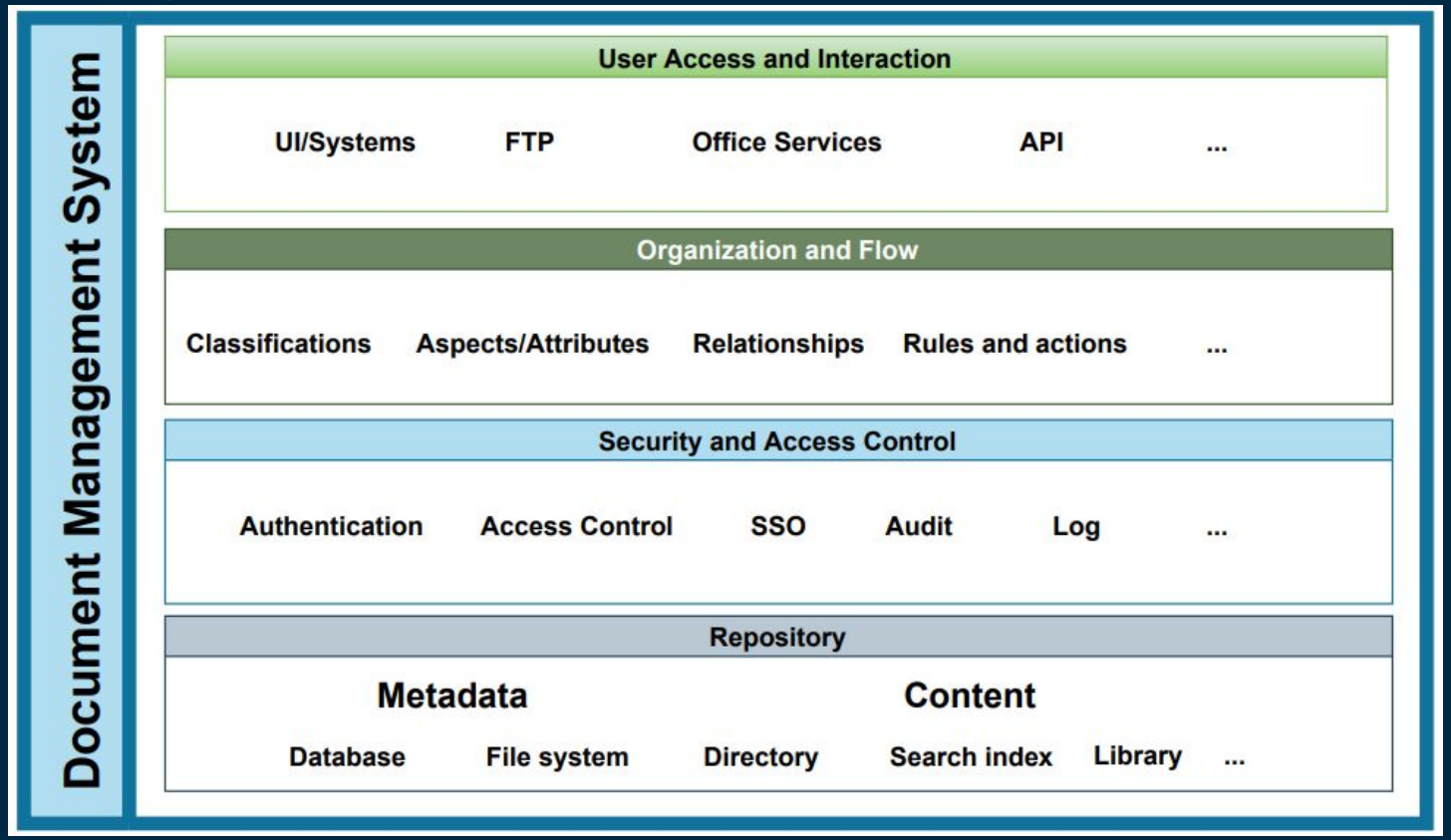
Scan->OCR->Correction

Digitalization for Office Automation



- A lot of time to scan documents
- Metadata recording using a tool
- Plan in advance – What is need for data science application?

Digitalization for Office Automation

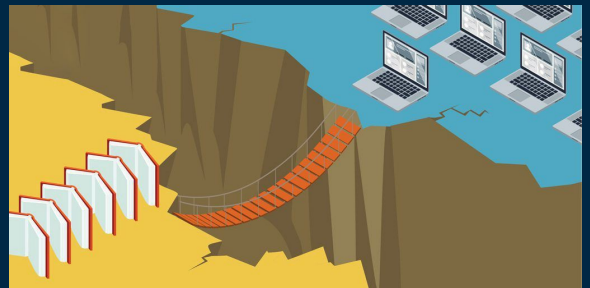


AI & Data Science Bottlenecks for Emerging Economies

- Lack of source data - Digitalization

- Lack of expertise - Developers, data scientist, ...

- Technology gap - Machines, IoTs, Storages



Top Areas of Digitalization for Data Science

- Healthcare – e-health

https://www.springboard.com/blog/wp-content/uploads/2020/11/shutterstock_1627208287-scaled.jpg

- Agriculture – smart-agriculture

https://media.springernature.com/lw685/springer-static/image/chp%3A10.1007%2F978-981-10-7578-0_5/MediaObjects/461002_1_En_5_Fig4_HTML.gif

- Government Offices – e-governance

https://akm-img-a-in.tosshub.com/businesstoday/images/story/201903/btsizeitsolutions_660_030719032604.jpg

- Finance/Banking – Fintech

https://fbf.eui.eu/wp-content/uploads/2022/02/fintech.jpg

- Education – e-learning

https://www.scnsoft.de/education-industry/elearning-portal/elearning-portals-cover-picture.svg

- Merchandise – e-commerce

https://retailingsa.co.za/wp-content/uploads/2021/04/ecommerce-business-blog-1.jpg

- Transportation

https://thesiliconreview.net/story_image_upload/article/thesiliconreview-big-data-analytics-in-public-transport.jpg

Data Science Initiatives in Africa

A great time to be on the African continent

Artificial Intelligence

Growth in the African Artificial Intelligence Community

Large push in research areas:

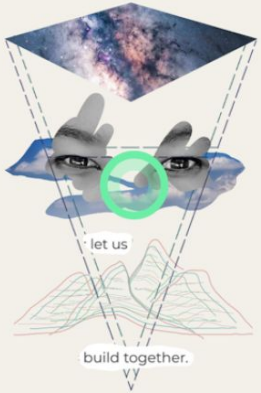
- Algorithms/Computation
- Climate Actions
- Economic Growth
- Education
- Inequalities
- Healthcare
- Industry, Innovation and Infrastructure
- African Language
- Sustainable Cities and Communities
- …

Driven by

- Grassroots AI
- Young People
- Open ML Tools
- Open Science
- Governments
- Private Sector

https://twitter.com/vukosi/status/1580080638916366336/photo/1

DSA
Data Science Africa



Masakhane

A grassroots NLP community for Africa, by Africans

DEEP LEARNING INDABA

DS-I Africa
Data Science for Health Discovery and Innovation in Africa

D
Data Science for Social Impact

Semantic Models for Ethiopic

AMHARIC NLP
BENCHMARK RESOURCES

Recommendations

- More Research on AI, Data Science, and ML projects
- Tailored for Data Science, NLP, AI related Master and Ph.D. streams
- More Conference and workshop, meetups, collocate with international events
- Training and consultation center for Data Science and AI projects
- Support startups in AI and Data Science

Thanks

Do you have any questions?

seid.muhie.yimam@uni-hamburg.de
+49 40 42883-2383
https://www.hcds.uni-hamburg.de/en/hcds/head-hcds/yimam.html

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik
Please keep this slide for attribution