

LEI deck

ACM SIGKDD 2022

Washington, DC Aug 13 - 18, 2022

KDD 2022 is a two-track conference hosting informative research and applied data science paper sessions. The conference is known for its highly selective, double-blind, peer-reviewed process. The program committee reviewed and accepted the following:

Track	Reviewed	Accepted	
Research Track Papers	1659	254	15.3%
Applied Data Science Track Papers	753	196	26%

14

Document Intelligence Workshop

<https://document-intelligence.github.io/DI-2022/>

ANDEA: Anomaly and Novelty Detection, Explanation, and Accommodation

<https://sites.google.com/view/andea2022/>

Online and Adaptive Recommender Systems (OARS)

<https://oars-workshop.github.io/>

Deep Search Relevance Ranking in Practice

Data science techniques for developing industry-scale search engines have long been a prominent part of most domains and their online products. Search relevance algorithms are key components of products across different fields such as e-commerce. In this tutorial, we plan to give an introduction to such large-scale search ranking systems, specifically focusing on deep learning methods. The topics will cover: (1) Overview of search ranking systems (2) Introduction to sequential and language models (3) Knowledge distillation approaches. For each of the sessions we plan to first give an introductory talk and then go over a hands-on tutorial. We will cover fundamental concepts using case studies, and hands-on examples, including the latest Deep Learning methods. Moreover, we plan to show example implementations of these methods in python, leveraging a variety of open-source machine-learning libraries as well as real industrial or open-source data.

<https://github.com/dlranking/dlrr>

15

Machine Learning in Finance

<https://sites.google.com/view/kdd-mlf-2022/schedule?authuser=0>

Decision Intelligence and Analytics for Online Marketplaces: Jobs, Ridesharing, Retail and Beyond

<https://sites.google.com/view/kdd22onlinemarketplaces>

Data Science and Artificial Intelligence for Responsible Recommendations

<https://rrs2022.github.io/Program/>

Workshop on Decision Intelligence and Analytics for Online Marketplaces: Jobs, Ridesharing, Retail, and Beyond

<https://sites.google.com/view/kdd22onlinemarketplaces>

ADS Invited: Sameena Shah (JP Morgan) and Vidhya Navalpakkam (Google) 16

ADS Invited: Sameena Shah (JP Morgan): Task Centric AI

The use of AI is gaining traction as organizations realize the advantages of using algorithms to streamline and improve the accuracy of tasks. In this talk, Sameena will build the case for task centric AI, using a variety of use cases from industry. Step through use cases that range from those that involve large amounts of data to no data, from detecting key insights to rare unexpected events, and from incorporating active domain expertise to passive sensor feedback.

ADS Invited: Vidhya Navalpakkam (Google): Accelerating eye movement research via ML-based smartphone gaze technology

Eye movements are thought to be a window to the mind, and have been extensively studied across Neuroscience, Psychology and HCI. However, progress in this area has been severely limited as the underlying eye tracking technology relies on specialized hardware that is expensive (upto \$30,000) and hard to scale. In this talk, I will present our recent work from Google, which shows that ML applied to smartphone selfie cameras can enable accurate gaze estimation, comparable to state-of-the-art hardware based mobile eye trackers, at 1/100th the cost and without any additional hardware. Via extensive experiments, we show that our smartphone gaze tech can successfully replicate key findings from prior eye movement research in Neuroscience and Psychology, across a variety of tasks including traditional oculomotor tasks, saliency analyses on natural images and reading comprehension. We also show that smartphone gaze could serve as a potential digital biomarker for detecting mental fatigue. These results show that smartphone gaze technology has the potential to unlock advances by scaling eye movement research, and enabling new applications for improved wellness and accessibility, such as gaze-based interaction for patients with ALS/stroke that cannot otherwise interact with devices

ADS Papers: Recommendation Systems & E-commerce 16

An Online Multi-task Learning Framework for Google Feed Ads Auction Models

<https://dl.acm.org/doi/abs/10.1145/3534678.3539055>

PM 3rd IADSS Workshop on Data Science Standards – Hiring, Assessing and Upskilling Data Science Talent

<https://www.iadss.org/kdd2022>

Model Monitoring in Practice: Lessons Learned and Open Challenges

With AI based models and solutions in high-stakes domains such as hiring, lending, criminal justice, healthcare, and education, it becomes critical to ensure that these models are making accurate predictions, are robust to shifts in the data, are not relying on spurious features, and are not unduly discriminating against minority groups. To this end, several approaches spanning various areas such as explainability, fairness, and robustness have been proposed in recent literature, and many papers and tutorials on these topics have been presented in recent computer science conferences. However, there is relatively less attention on the need for monitoring machine learning (ML) models once they are deployed and the associated research challenges. In this tutorial, we first motivate the need for ML model monitoring, as part of a broader AI model governance and responsible AI framework, and provide a roadmap for thinking about model monitoring in practice. We then present findings and insights on model monitoring desiderata based on interviews with various ML practitioners. We then describe the technical considerations and challenges associated with realizing the above desiderata in practice. We provide an overview of techniques/tools for model monitoring. Then, we focus on the real-world application of model monitoring methods and tools, present practical challenges/guidelines for using such techniques effectively, and lessons learned from deploying model monitoring tools for several web-scale AI/ML applications. We present case studies across different companies, spanning application domains such as financial services, healthcare, hiring, conversational assistants, online retail, computational advertising, search and recommendation systems, and fraud detection. We hope that our tutorial will inform both researchers and practitioners, stimulate further research on model monitoring, and pave the way for building more reliable ML models and monitoring tools in the future.

<https://sites.google.com/view/model-monitoring-tutorial>

Towards Universal Sequence Representation Learning for Recommender Systems

<https://arxiv.org/abs/2206.05941>

Temporal Graph Learning for Financial World: Algorithms, Scalability, Explainability & Fairness

The most intuitive way to model a transaction in the financial world is through a Graph. Every transaction can be considered as an edge between two vertices, one of which is the paying party and another is the receiving party. Properties of these nodes and edges directly map to business problems in the financial world. The problem of detecting a fraudulent transaction can be considered as a property of the edge. The problem of money laundering can be considered as a path-detection in the Graph. The problem of a merchant going delinquent can be considered as the property of a node. While there are many such examples, the above help in realising the direct mapping of Graph properties with the financial problems in the real-world.

This tutorial will build upon this promise of using Graph based Learning for solving business problems of the financial world. We will start with representing transactions as a Graph. We will then get into the specifics of the transaction graph and highlight the challenge with respect to the transactions graph. These challenge are related to (a) the size of the graph (over 100 million nodes and billions of edges), (b) the temporal nature of the graph (new transactions keep appearing every second, new merchants and cards keep appearing every day - thus the graph is always evolving), and, (c) financial world being a heavily regulated industry, the questions around interpretability and fairness are critical to them being of any practical use.

After setting the business context, this tutorial will present latest developments in machine learning from large graphs. We will show how some of our work (BipGNN) and that in the research community (DGL) has scaled to graphs of over billion edges and 100 million nodes. We will then present the research in Temporal Graph Networks to handle dynamic graphs. Finally, we will highlight the challenges with respect to the bias in Graph based algorithms and discuss ways to mitigate it.

ADS Paper Showcase: Conversation, QA and Other NLP Applications 18

DocLayNet: A Large Human-annotated Dataset for Document Layout Segmentation

<https://arxiv.org/abs/2206.01062>

Ask to know more: Counterfactual Explanations for Fake Claims

<https://arxiv.org/pdf/2206.04869.pdf>

ADS Paper Showcase: Recommendation & Contextualization 18

Surrogate for Long-Term User Experience in Recommender Systems

<https://dl.acm.org/doi/abs/10.1145/3534678.3539073>

Automatically Discovering User Consumption Intents in Meituan

<https://dl.acm.org/doi/abs/10.1145/3534678.3539122>

PinnerFormer: Sequence Modeling for User Representation at Pinterest

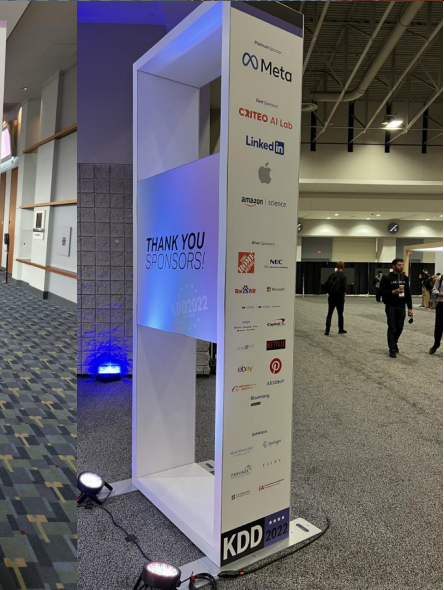
<https://arxiv.org/pdf/2205.04507.pdf>

ItemSage: Learning Product Embeddings for Shopping Recommendations at Pinterest

<https://arxiv.org/pdf/2205.11728.pdf>

NxtPost: User to Post Recommendations in Facebook Groups

<https://arxiv.org/pdf/2202.03645.pdf>



BALLROOM ABC EVENT SCHEDULE

MONDAY, AUGUST 15
5:15 PM - 6:45 PM
KDD 2022 Opening Session
Welcome Remarks, SIGKDD Awards, KDD 2022 Stats & Highlights

TUESDAY, AUGUST 16
8:00 AM - 8:15 AM
Welcome and Announcements
Opening Keynote Address
The Power of Statistical/Relational Thinking by Lise Getoor

WEDNESDAY, AUGUST 17
8:00 AM - 8:15 AM
Welcome and Announcements
Keynote Address
AI for social impact: Results from deployments for public health and conversation by Milind Tambe

6:00 PM - 9:00 PM
KDD 2022 Celebration Dinner
sponsored by Meta

THURSDAY, AUGUST 18
8:00 AM - 8:15 AM
Welcome and Announcements
Keynote Address
Beyond Traditional Characterizations in the Age of Data: Big Models, Scalable Algorithms, and Meaningful Leadership by Cheng-Nan Ting

3:30 PM - 4:00 PM
KDD Closing Ceremony





References

GradMask: Gradient-Guided Token Masking for Textual Adversarial Example Detection

<https://openreview.net/forum?id=flmoryxnH0Z>

Personalized Chit-Chat Generation for Recommendation Using External Chat Corpora

https://www.microsoft.com/en-us/research/uploads/prod/2022/05/Conversation_generation_for_news_recommendation_camera_ready.pdf

GDMix: A deep ranking personalization framework

<https://engineering.linkedin.com/blog/2020/gdmix--a-deep-ranking-personalization-framework>

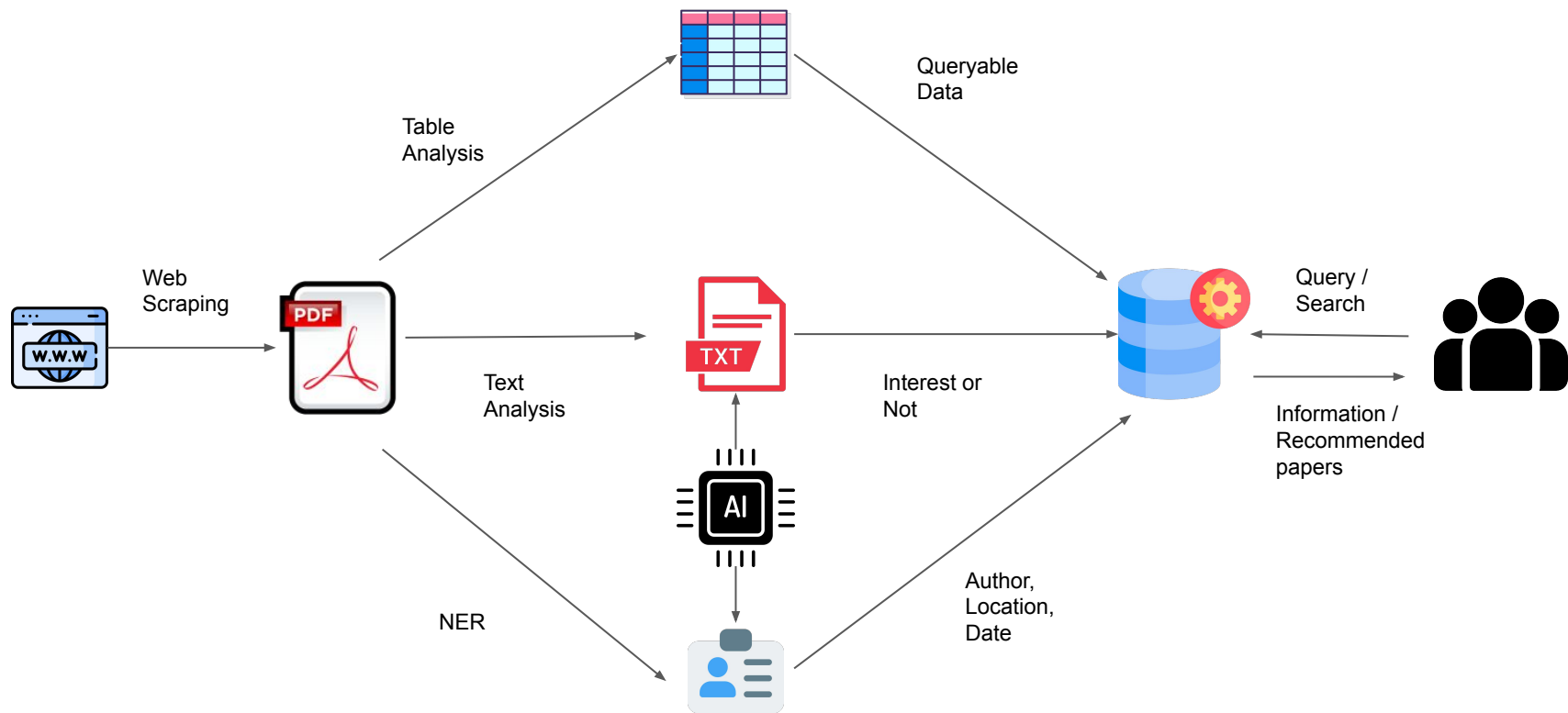
**KDD 2022 Workshop on Data Science and Artificial Intelligence
for Responsible Recommendations (DS4RRS)**

<https://rrs2022.github.io/Program/>

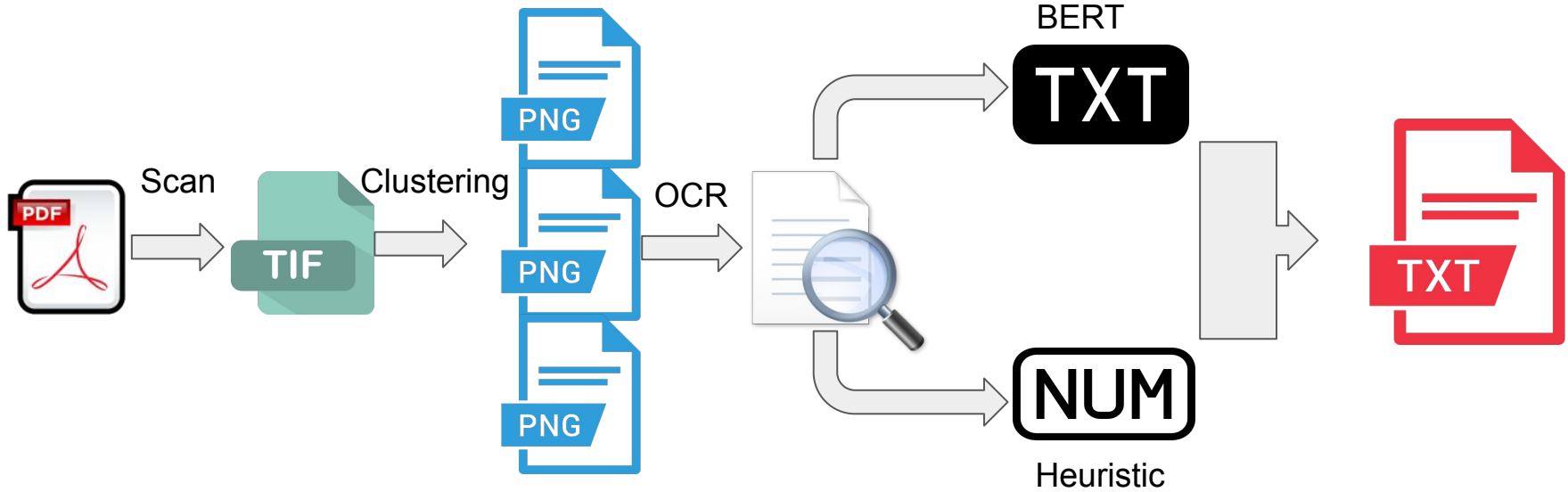
Algorithmic Fairness on Graphs Methods and Trends

http://jiank2.web.illinois.edu/tutorial/kdd22/algofair_on_graphs.html

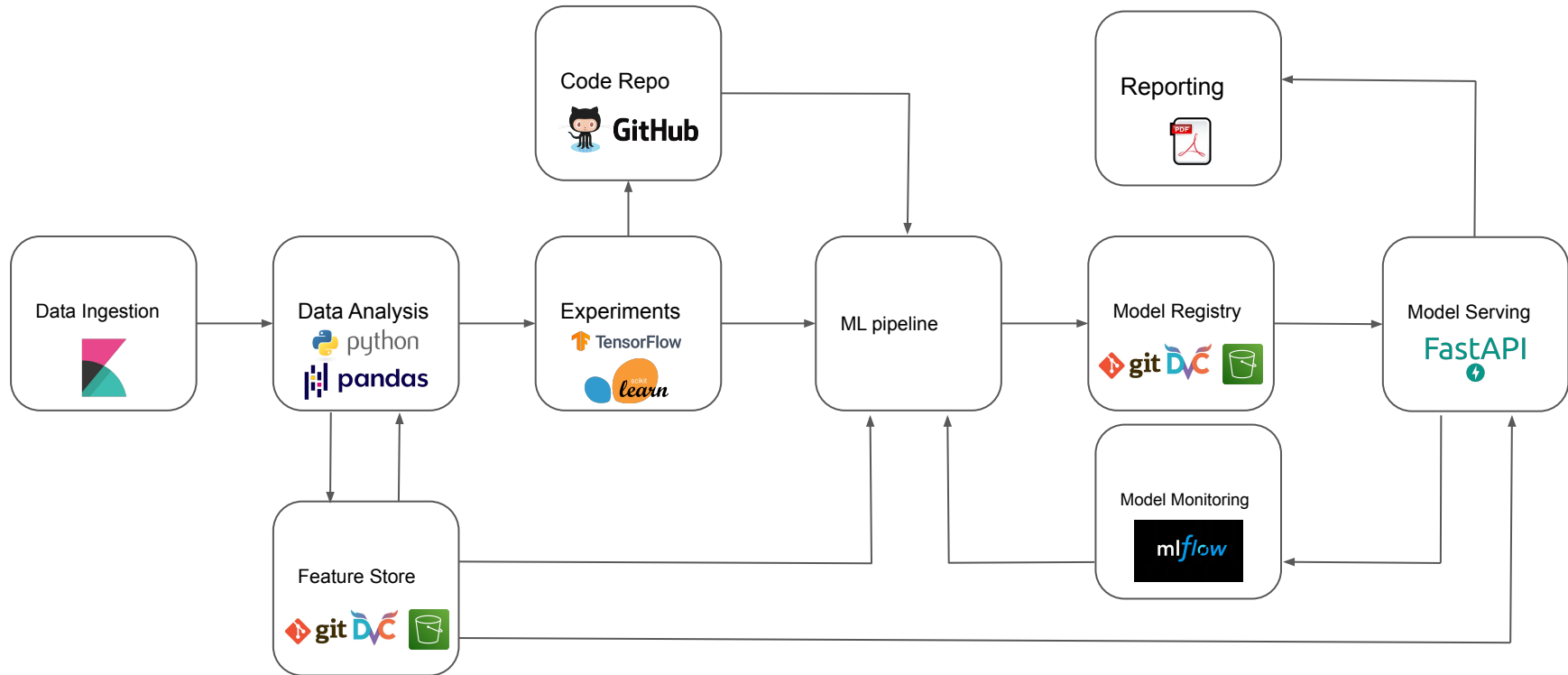
System Overview



Seaworthy Architecture

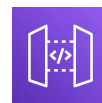


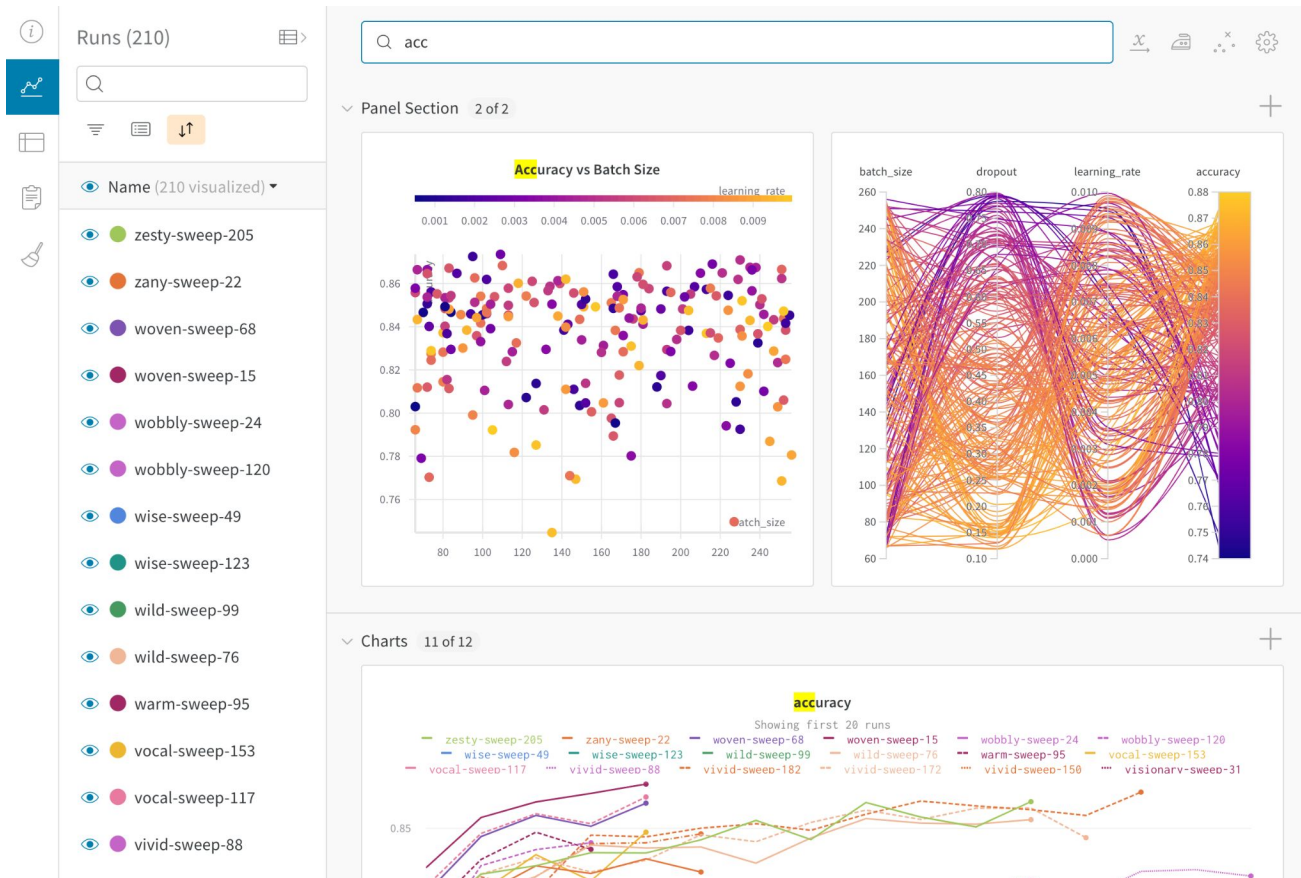
3. ML pipeline



3. ML pipeline







<https://wandb.ai/site/pricing>

<https://hydra.cc/docs/intro#community>

Create

Frameworks



Native support

Converters

Native support

ONNX Model

Deploy

Azure

Azure Machine Learning services

Ubuntu VM

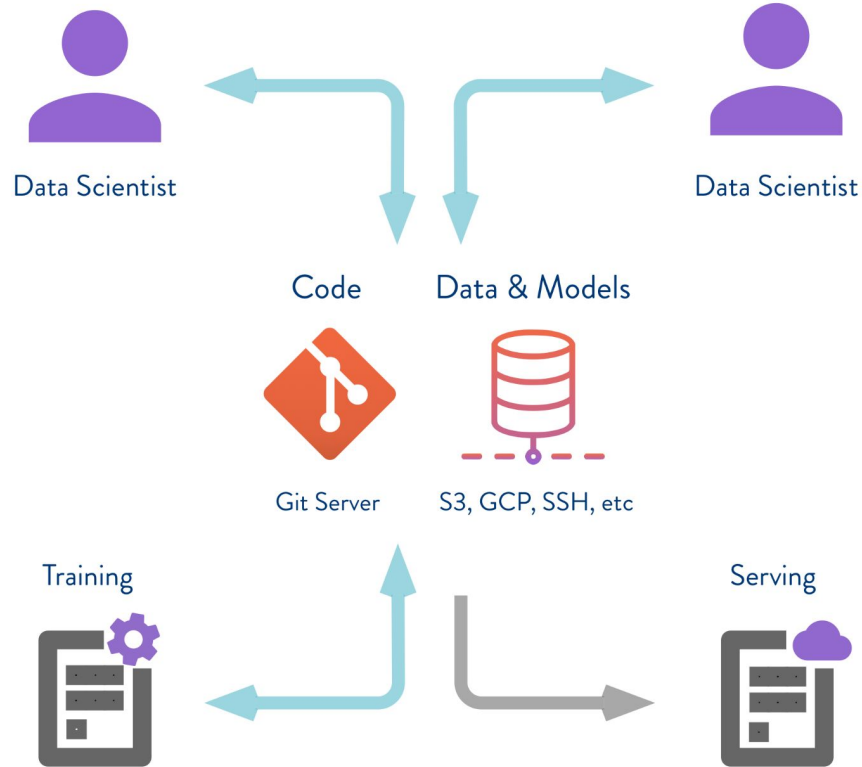
Windows Server 2019 VM

Native support

Windows Devices

Other Devices
(iOS, etc)

Converters

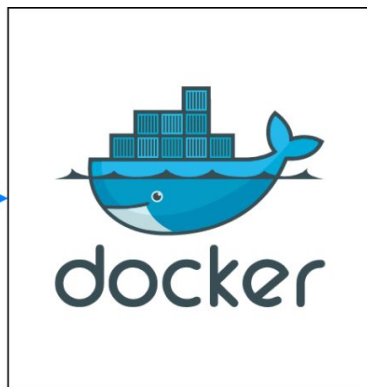


<https://dvc.org/doc/install/macos>



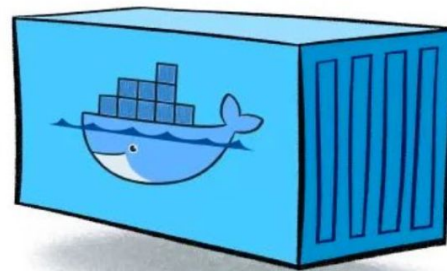
DockerFile

BUILD

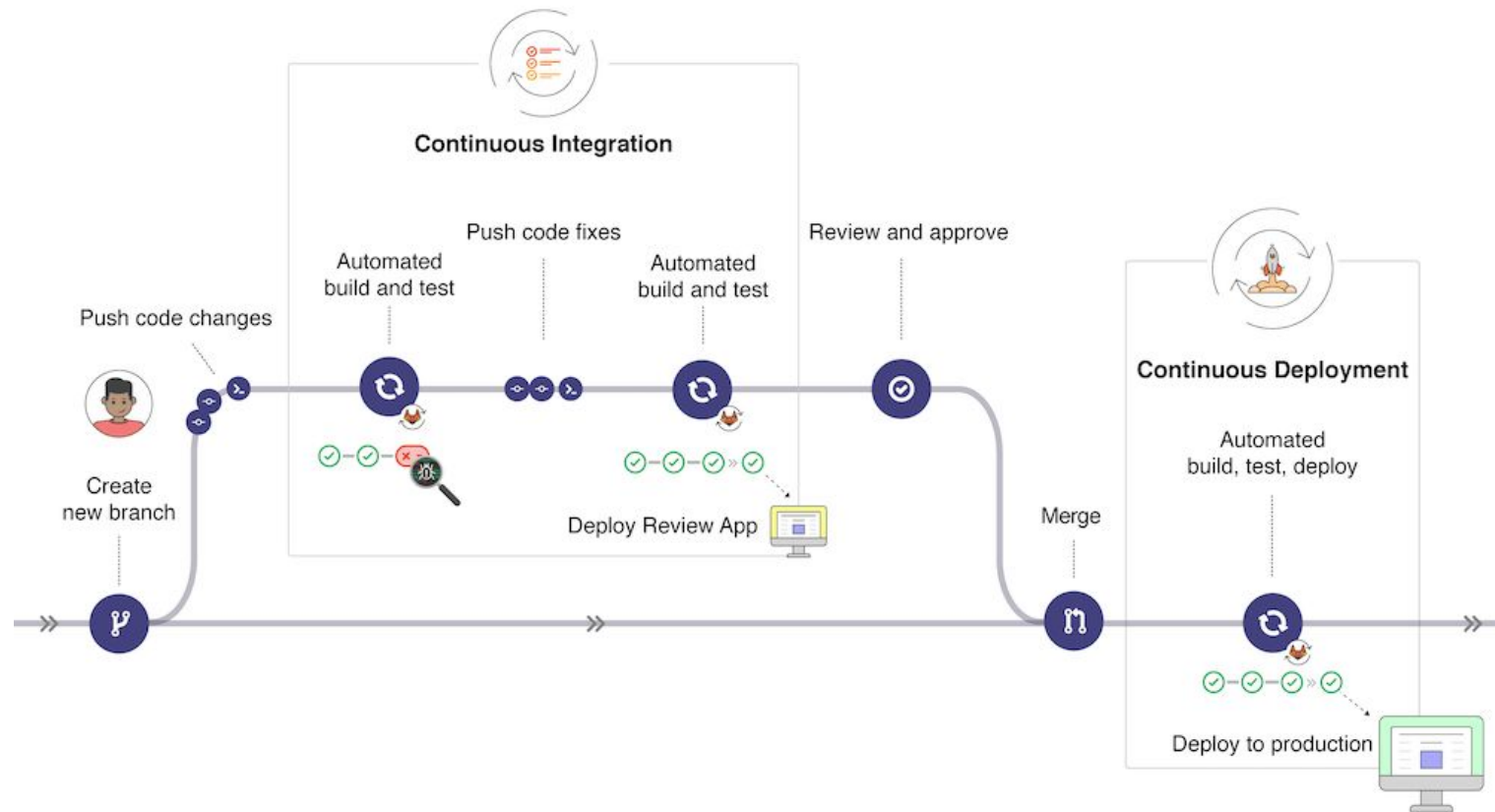


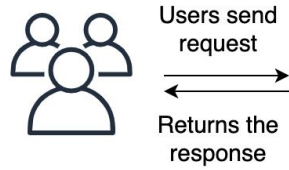
Docker Image

RUN

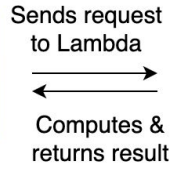


Docker Containers

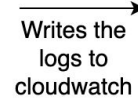




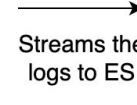
API
Gateway



Lambda



Cloudwatch
Logs



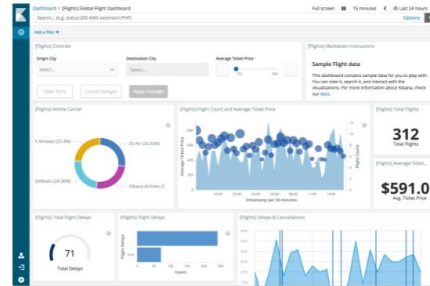
Elasticsearch
Cluster

Fetches the logs from ES



Kibana

Updates the dashboard with latest logs



Dashboard

BDSI

BANK DATA SCIENCE INNOVATION