



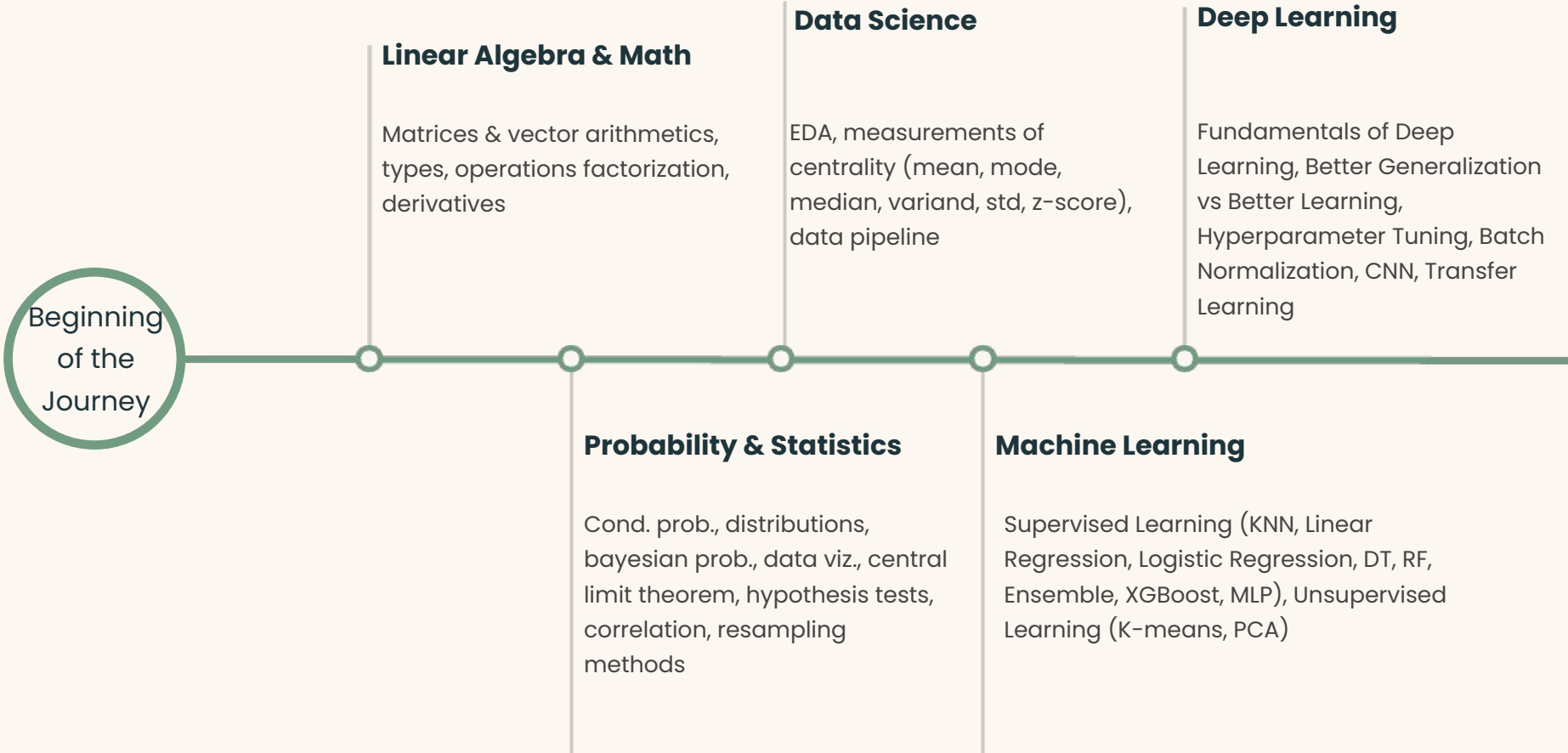
DCA0305

ivanovitch.silva@ufrn.br

# Machine Learning Based Systems Design

---

Beginning  
of the  
Journey



## Linear Algebra & Math

Matrices & vector arithmetics,  
types, operations factorization,  
derivatives

## Probability & Statistics

Cond. prob., distributions,  
bayesian prob., data viz., central  
limit theorem, hypothesis tests,  
correlation, resampling  
methods

## Data Science

EDA, measurements of  
centrality (mean, mode,  
median, variand, std, z-score),  
data pipeline

## Machine Learning

Supervised Learning (KNN, Linear  
Regression, Logistic Regression, DT, RF,  
Ensemble, XGBoost, MLP), Unsupervised  
Learning (K-means, PCA)

## Deep Learning

Fundamentals of Deep  
Learning, Better Generalization  
vs Better Learning,  
Hyperparameter Tuning, Batch  
Normalization, CNN, Transfer  
Learning



## **TinyML**

Optimization,  
quantization, deploy  
into a microcontroller

## **New Models**

Transformers, Difusion, GAN,  
LLM, GNN, Generative Flow  
Networks

## **But ....**

If a model is not  
deployed, it does not  
generate value

Cont.

2016

Pandas

2017

Scikit-Learn

2018-2020

Keras, Pytorch, Covid-19

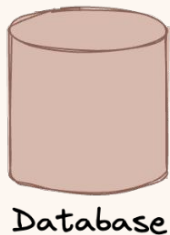
2021-2023

MLOps, GAN, LLM

How to project a  
typical machine  
learning workflow?



What is the issue with this solution?



```
clf = xgboostClassifier()
```

```
clf.fit(train_x, train_y)
```

```
pred = clf.predict(test_x)
```

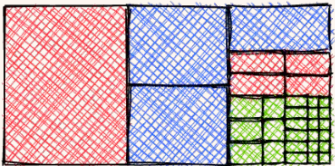
```
some_metrics(test_y, pred)
```



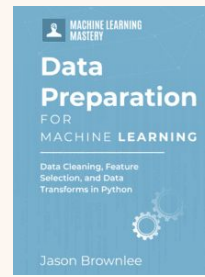
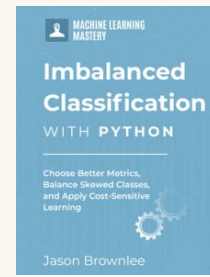
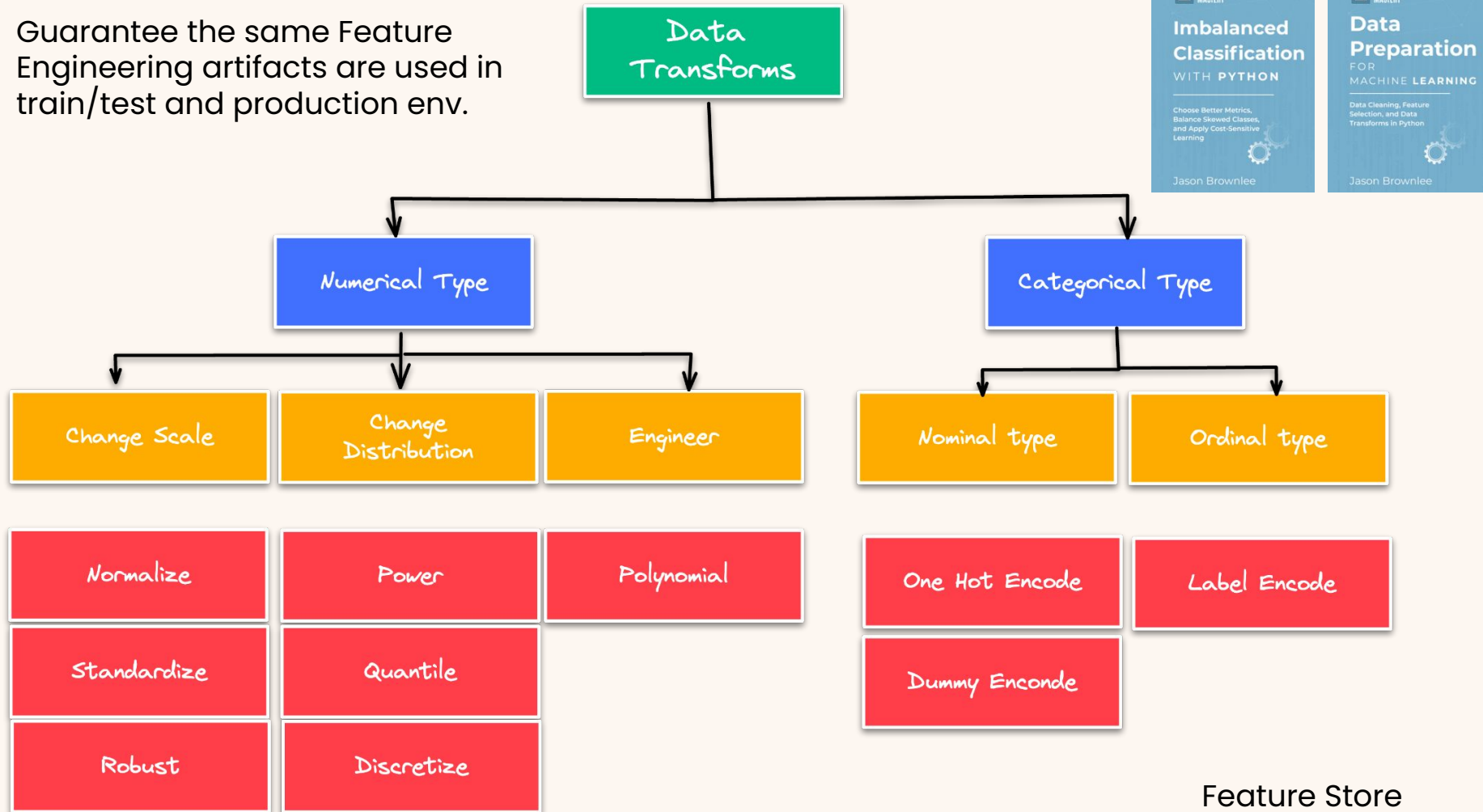
publish a journal/conference paper, report



Stakeholders






Guarantee the same Feature Engineering artifacts are used in train/test and production env.



Feature Store

## Article

# Predictive Models for Imbalanced Data: A School Dropout Perspective

Thiago M. Barros <sup>1,\*</sup>, Plácido A. Souza Neto <sup>1,†</sup> and Ivanovitch Silva <sup>2,†</sup>  
and Luiz Affonso Guedes <sup>2,†</sup>

<sup>1</sup> Federal Institute of Rio Grande do Norte (IFRN), 1559 Tirol Natal, Brazil; placido.neto@ifrn.edu.br

<sup>2</sup> Federal University of Rio Grande do Norte (UFRN), 59078-970 Natal, Brazil; ivan@imd.ufrn.br (I.S.);  
affonso@dca.ufrn.br (L.A.G.)

\* Correspondence: thiago.medeiros@ifrn.edu.br

† These authors contributed equally to this work.

Received: 24 July 2019; Accepted: 10 November 2019; Published: 15 November 2019



**Abstract:** Predicting school dropout rates is an important issue for the smooth execution of an educational system. This problem is solved by classifying students into two classes using educational activities related statistical datasets. One of the classes must identify the students who have the tendency to persist. The other class must identify the students who have the tendency to dropout. This problem often encounters a phenomenon that masks out the obtain into this phenomenon and provides a reliable educational data mining predicts the dropout rates. In particular, the three data classifying techniques neural networks and Balanced Bagging, are used. The performances of the models with and without the use of a downsample, SMOTE and ADASYN data balancing other parameters geometric mean and UAR provides reliable results with dropout rates using Balanced Bagging classifying techniques.

**Keywords:** dropout rates; accuracy paradox; imbalanced learning; decision tree; mlp; decision tree; Balanced Bagging; UAR; SMOTE; ADASYN

Concept/  
Data Drift

## Evasão escolar de crianças e adolescente aumenta 171% na pandemia, diz estudo

Levantamento da organização Todos Pela Educação mostra que 244 mil crianças de 6 a 14 anos estavam fora da escola no segundo trimestre de 2021.

Por g1 — São Paulo  
02/12/2021 13h28 · Atualizado há um ano



## Pandemia aumenta evasão escolar, diz relatório do Unicef

A quantidade de alunos, com idades entre 6 e 17 anos, que abandonaram as instituições de ensino foi de 1,38 milhão



## Educadores alertam para aumento de evasão escolar durante a pandemia

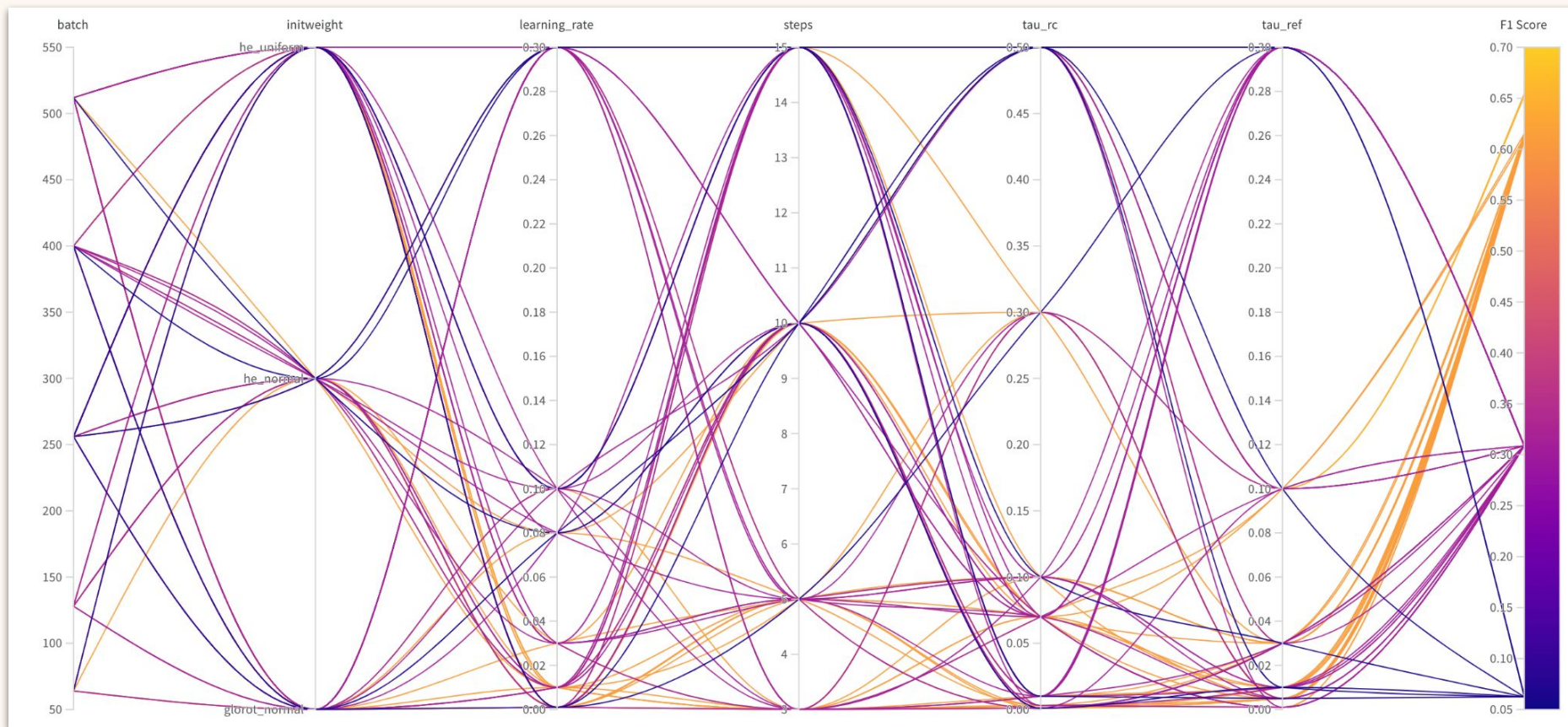
Para debatedores, desafio agora é atrair estudantes de volta à escola e recuperar o aprendizado



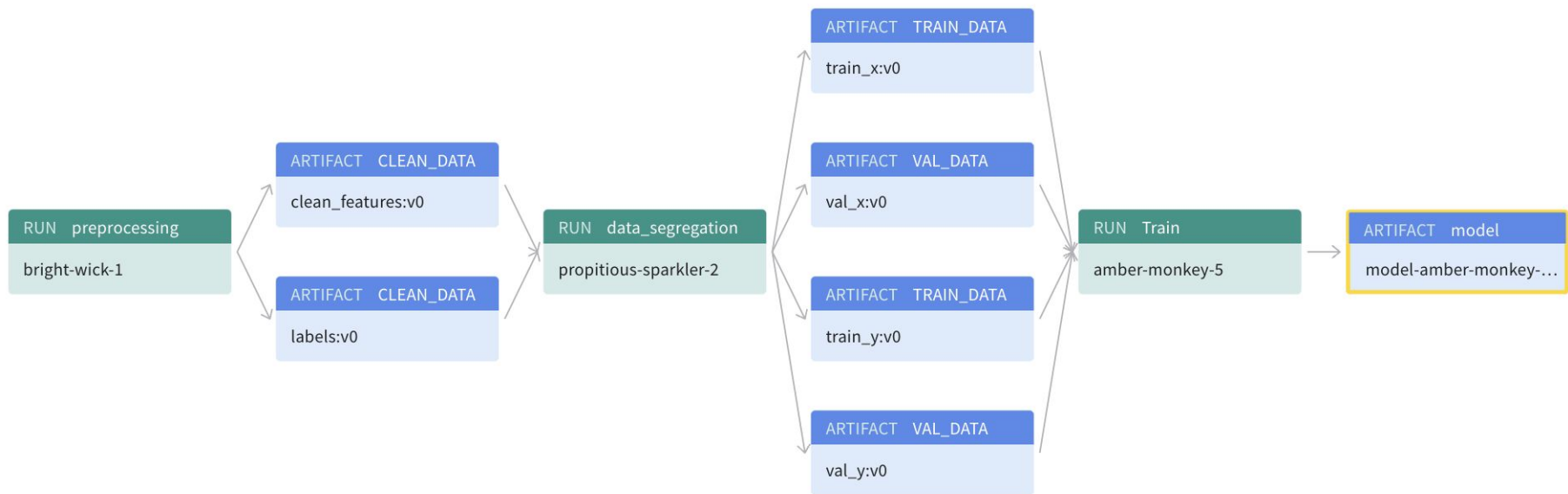
06/10/2021 - 20:04

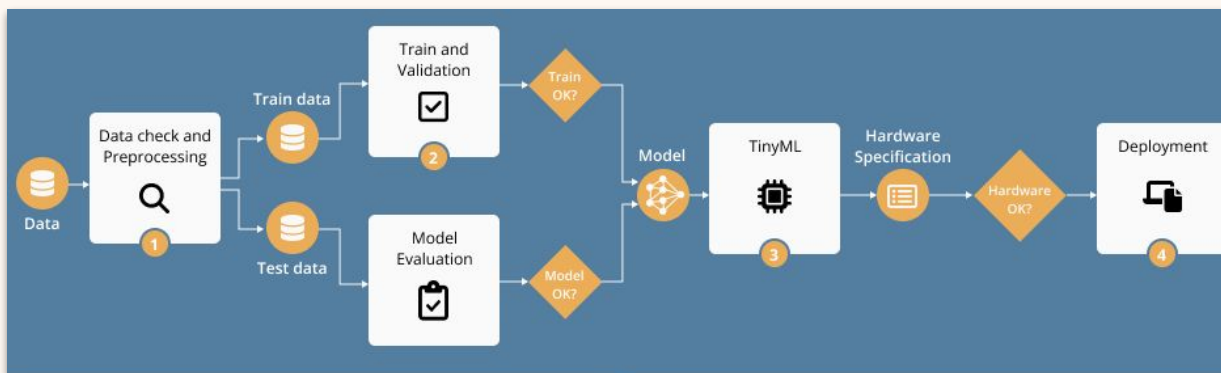


# What settings were used in the last experiment?

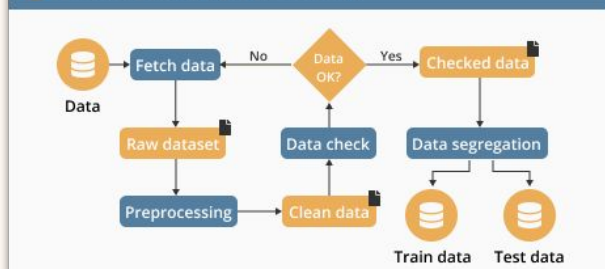


# A more efficient machine learning workflow

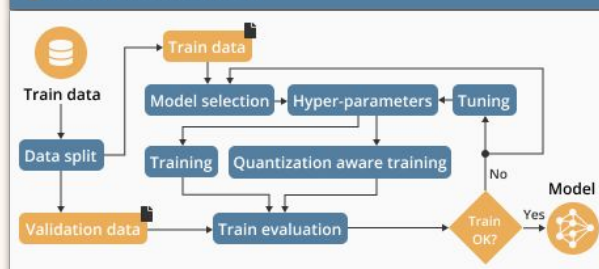




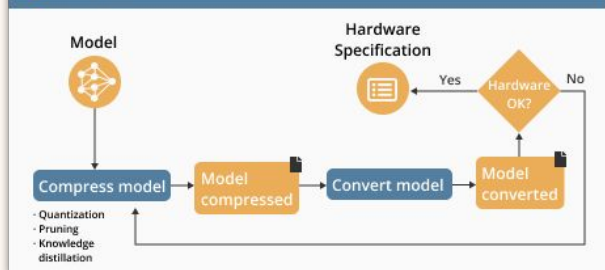
### 1 Data check and preprocessing



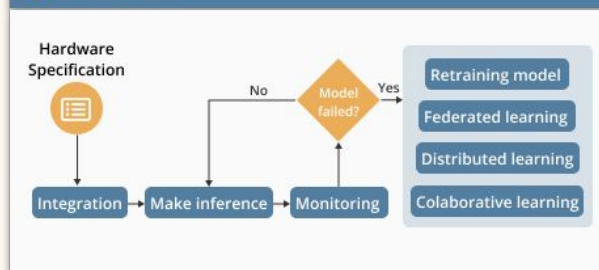
### 2 Train and validation



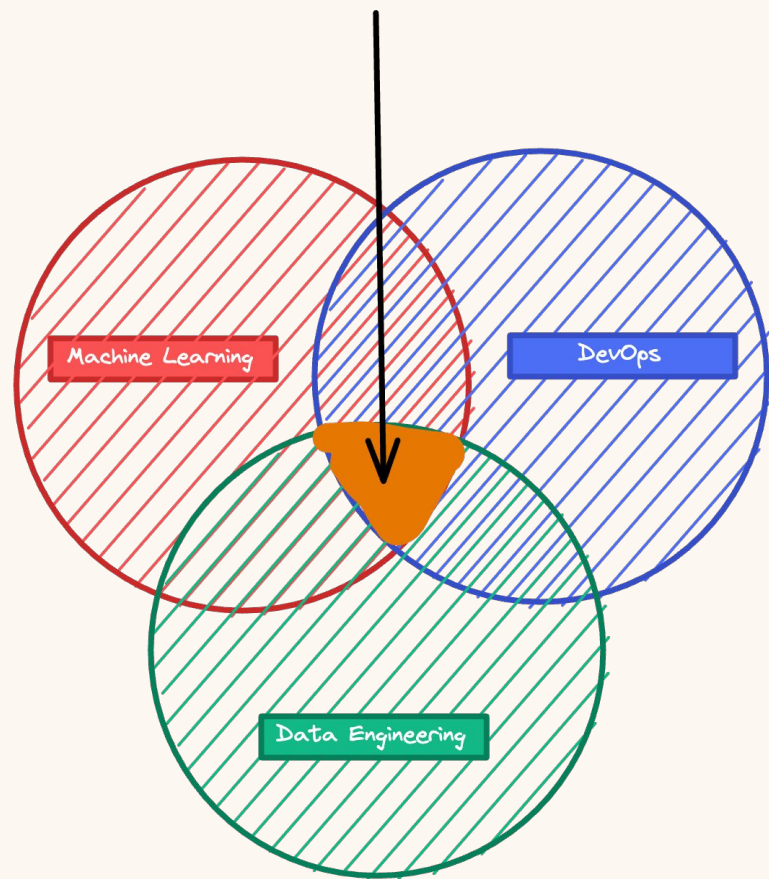
### 3 TinyML



### 4 Deployment



# MLOps

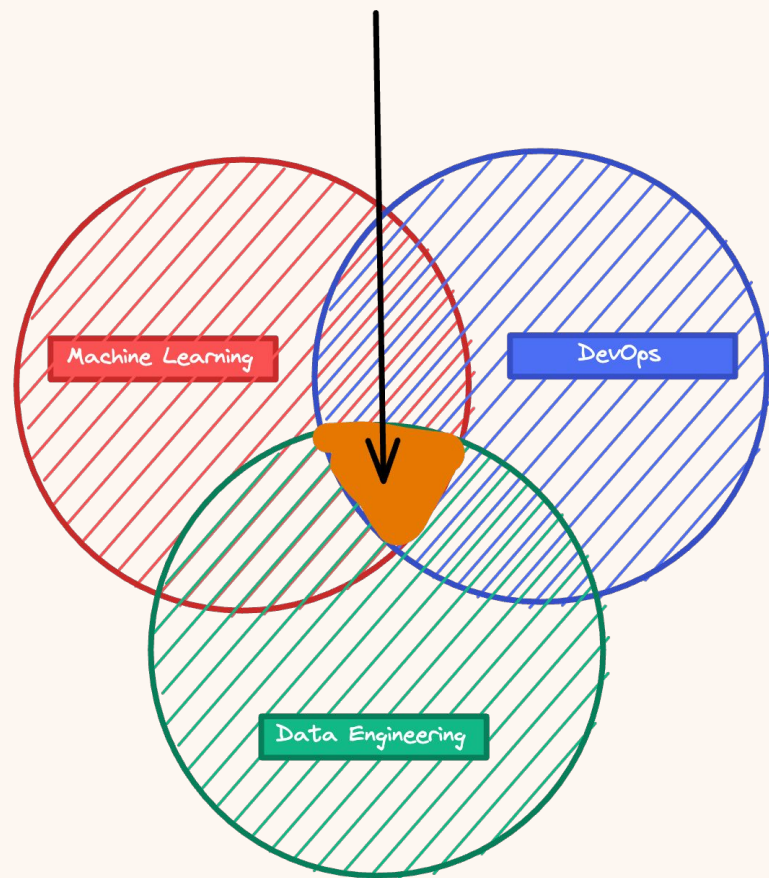


## Python Essentials for MLOps

CLI Fundamentals	Clean Code Principles	Production Ready Code	Programming
Elements of the Command Line	Refactoring Documentation	Catching Errors Logging	Functions Classes Decorators
Infrastructure Github Codespace vscode, colab, terminal	Python Code Quality Authority (PCQA)	Testing	Interact with APIs and SDKs to build command-line tools



# MLOps



## Essential guide for NLP

Foundations

Data  
Preparation

Applications

ML  
Introduction

From  
Linguistics  
to NLP

Tokenization  
Bag of  
Words

Text  
Classification

Foundations  
of Attention

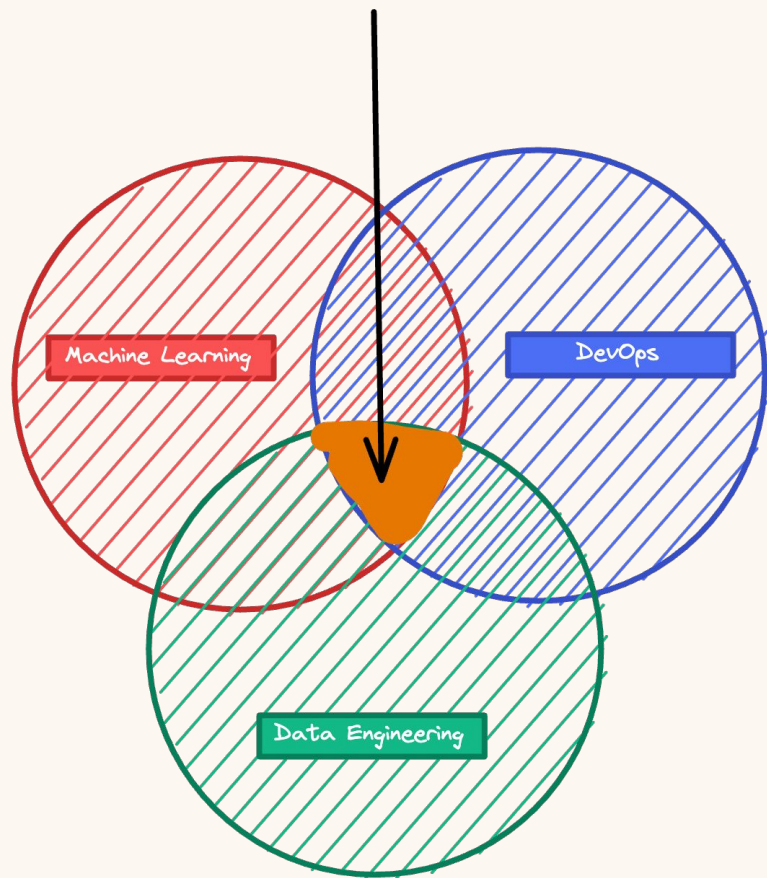
Standard  
Models

Word  
Embeddings

Image  
Captioning  
Machine  
Translation

From RNN  
to  
Transformer

# MLOps



## Building a Reproducible Model Workflow

Principles  
and  
Practices of  
MLOps

Full-  
Pipeline

MLOps  
Tools

Other  
Tools

Operations  
Pipelines

ETL, Data  
Segregation,  
Data Checks,  
Train, Validation  
and Experiment  
Tracking

Introduction  
to MLflow  
and  
HuggingFace

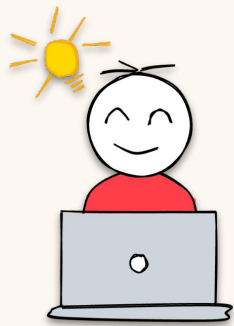
FastAPI,  
DVC,  
Gradio,  
Streamlit

End to End  
MLOps  
and AIOps

Data and  
Artifact Version,  
Release for  
reproducibility,  
Deploy

Deploying  
and Applied  
HuggingFace

Weights &  
Bias



## Work in Progress WiP



Third\* time the course is offered

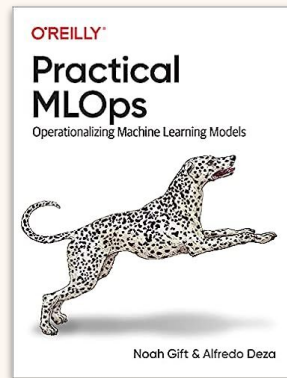


The subject is new, we don't have all the answers  
We are all learning together!!!

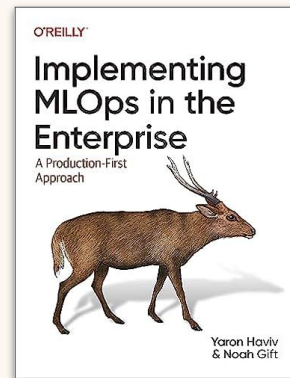


We appreciate you:

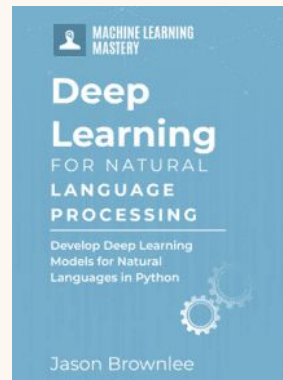
- a) **enthusiasms** for trying out new things
- b) **patience** bearing with things that don't quite work
- c) **feedback** to improve the course



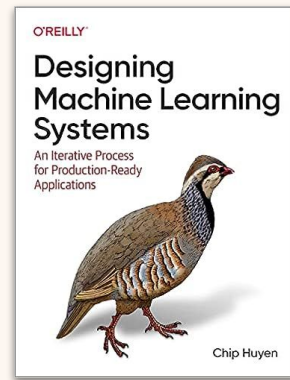
Sept. 2021



Oct. 2023

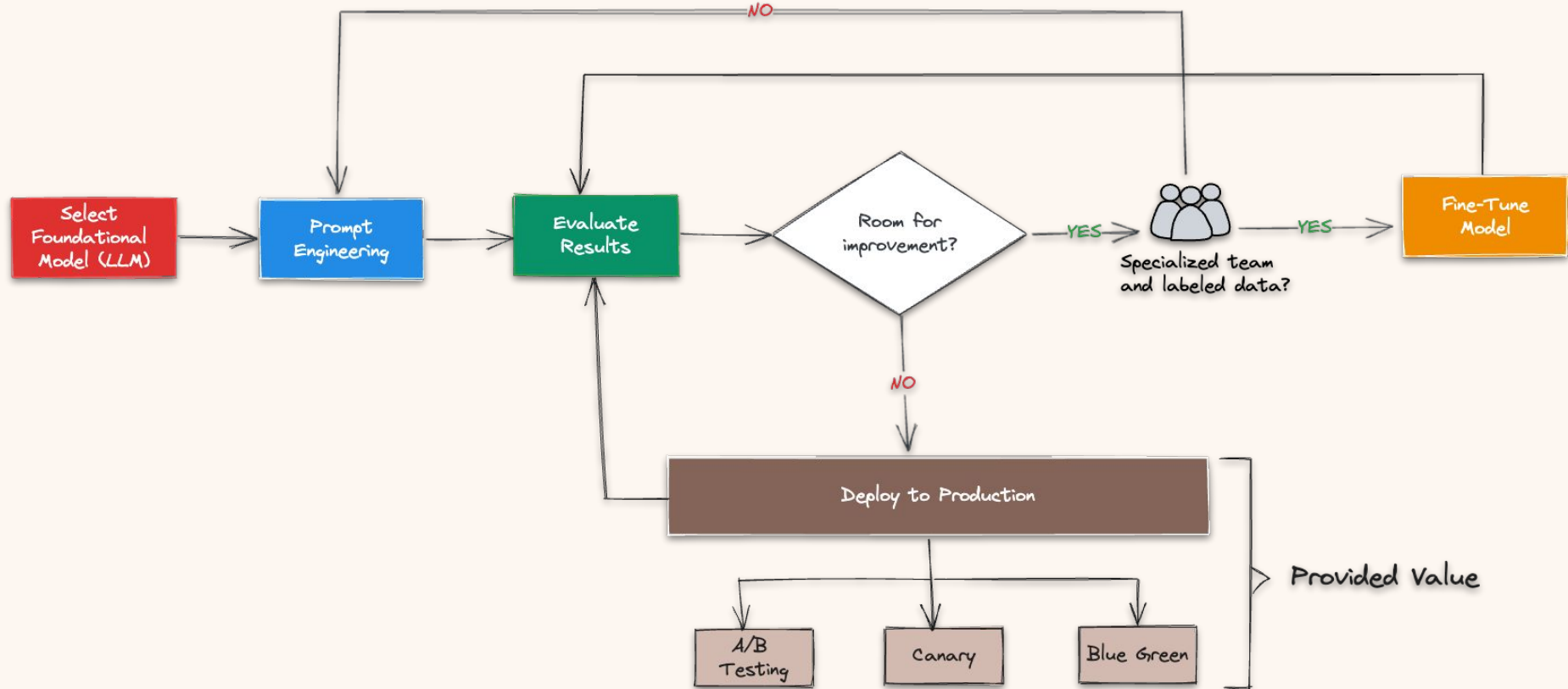


March 2021



May 2022

# Large Language Model Ops (LLMOps)





# Clone me!!!!

<https://github.com/ivanovitchm/mlops>