

Machine learning workshop with TensorFlow

Vijayabharathi Karuppasamy



ImtheVB

About me

- Profession - Agile, DevOps implementation, Automation
- Passion - Machine learning
- Roles - Google India ML developer Lead, ML facilitator
- Love - Eat, Code, Travel and meeting people
- Work - Technology Analyst at Infosys
- Part of - Google Cloud Developer community Chennai, GDG
- Tech enthusiast, Mechanical Engineer

What we are going to do ?

- Machine learning and Data science
- What Industry is doing in ML?
- ML using Python in Jupyter notebook & Google colab
- ML using TensorFlow (Hands-on and Demos)
- ML learning path, career path, Best practices
- Questionnaire, resources and feedback

Machine learning and Data science

- What is Machine learning ?
- Analyzing the problem for ML
- AI vs Machine learning vs Deep Learning
- Machine learning processes
- What is Data science?
- Basic terminologies in Data science
- Why Python ? Why not R ?
- Python libraries

Machine learning and Data science

- Data processing and visualization
- Various algorithms and selection of algorithms
- Model evaluation and boosting
- Improve model training, Feature engineering
- Deployment and Maintenance
- Myths of ML and AI
- Points to be remembered while doing ML

What Industry is doing in ML?

- How to use ML in production environment (Sample chat bot)
- Google, Twitter, Hike, Amazon use cases
- Current research areas
- Build for digital India program
- AI research lab by Google in India
- ML dev leads summit
- Auto ML and ML.NET

ML using Python in Jupyter notebook & Google colab

- CPU, GPU, TPU and Edge computing
- Jupyter notebook
- Google colab
- Free TPUs for ML research
- Sample Machine learning problem in Colab

ML using TensorFlow

- TensorFlow libraries – Classification hands-on
- What is new in TensorFlow 2.0
- TensorFlow Lite
- TensorFlow for JS - Demo
- TensorFlow Extended - workshop
- TensorFlow for Swift

ML learning path, career path, Best practices

- Where to start ? How to start ?
- Various learning resources available
 - Books
 - Journals
 - Blogs & documentation
 - Datasets
 - Courses and on-line references
- Full stack machine learning engineer
- Career path

Let's see more in workshop...!

What is Machine learning ?

- Machine + learning = Machine learning
- Predicting future
- Organizing or grouping
- Full of Mathematics!
- Full of Python code!
- I don't know...!

What is Machine Learning?

- Data Science technique, which allows the computers to use the existing data to forecast the future trends.
- Branch of computer science in which a computer "learns" from data in order to perform predictive analytics
- Machine Learning finds patterns in large volumes of data and uses those patterns to perform predictive analysis
- Supervised learning
 - Regression and classification
- Unsupervised learning
 - Clustering



Why Machine learning?

THE DATA SCIENCE / ANALYTICS LANDSCAPE



2,350,000

DSA job listings in 2015

By 2020, DSA job openings
are projected to grow

15%

364,000

Additional job listings
projected in 2020

Demand for both Data
Scientists and Data Engineers
is projected to grow

39%

DSA jobs remain open

5 days

longer than average

DSA jobs advertise average salaries of

\$80,265

With a premium over all BA+ jobs of

\$8,736

81%

Of DSA jobs require workers with
3-5 years of experience or more

AI vs Machine learning vs Deep Learning

Artificial Intelligence

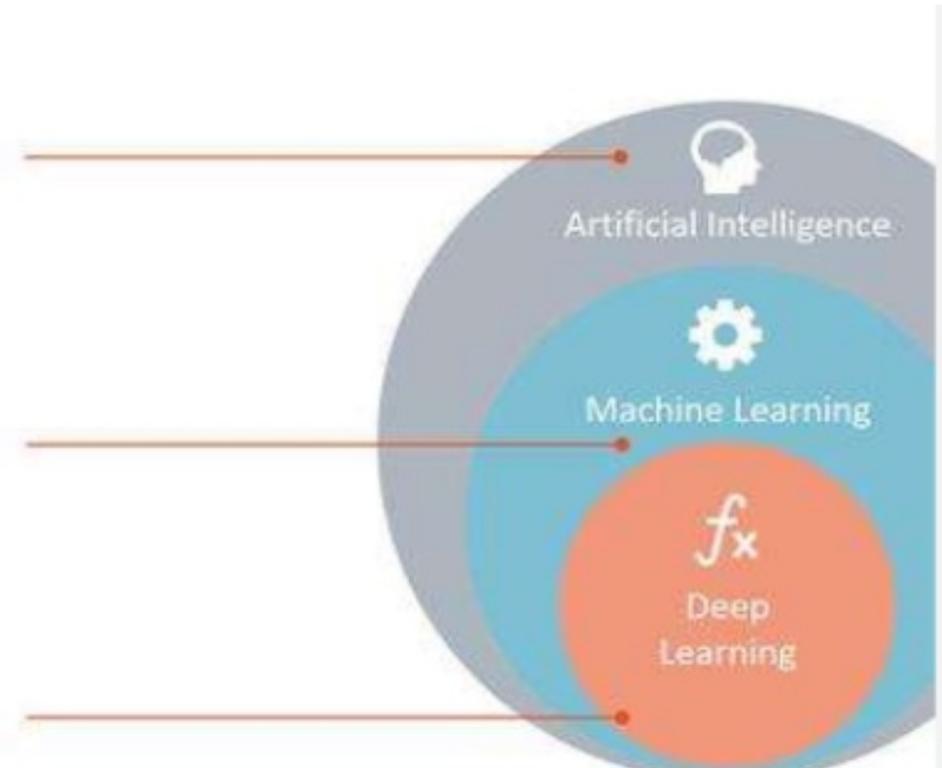
Any technique which enables computers to mimic human behavior.

Machine Learning

Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

Deep Learning

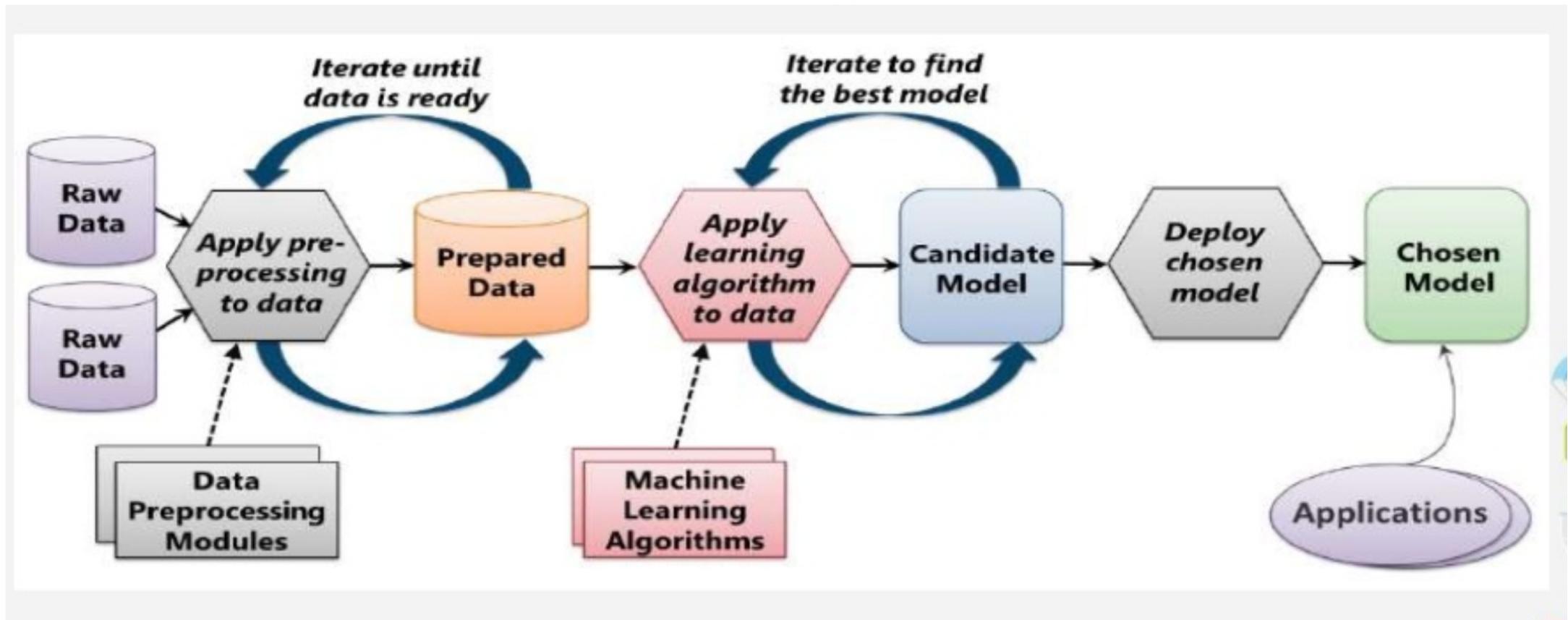
Subset of ML which make the computation of multi-layer neural networks feasible.



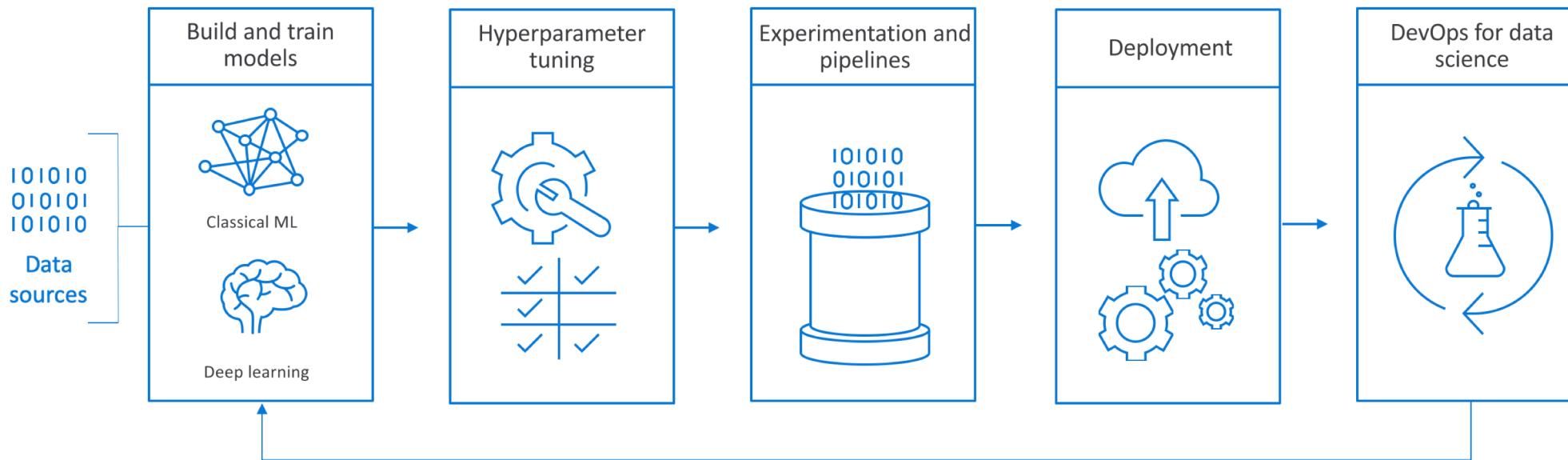
Analyzing the problem for ML

- Is it AI problem?
- Can it be solved with AI?
- Do we have enough data?
- Do we have labelled data?
- Checking orientation from image or recreating face from audio?

Machine learning processes



What is Data science?



Basic terminologies in Data science

- Features
- Model
- Target
- Labels, Classes
- Dataset
- Train, Test data
- Under Fit, Over Fit
- Accuracy, Precision, RMSValue, error etc..etc..

Data processing and visualization

- **Data Validation**

- Is Data Relevant?
 - Domain Knowledge
 - Correlation vs Causation
- Is data Variance?
- Is data Enough?
- Imbalanced Data?

- **Data Clean-up**

- Access Data
- Clean-up
 - Remove missing values or fill missing values
- Outlier Treatment

- **Visualization**

- Histogram
- ScatterPlot
- PairPlot
- Boxplot, ViolinPlot
- Correlation Plot

Why Python ? Why not R ?

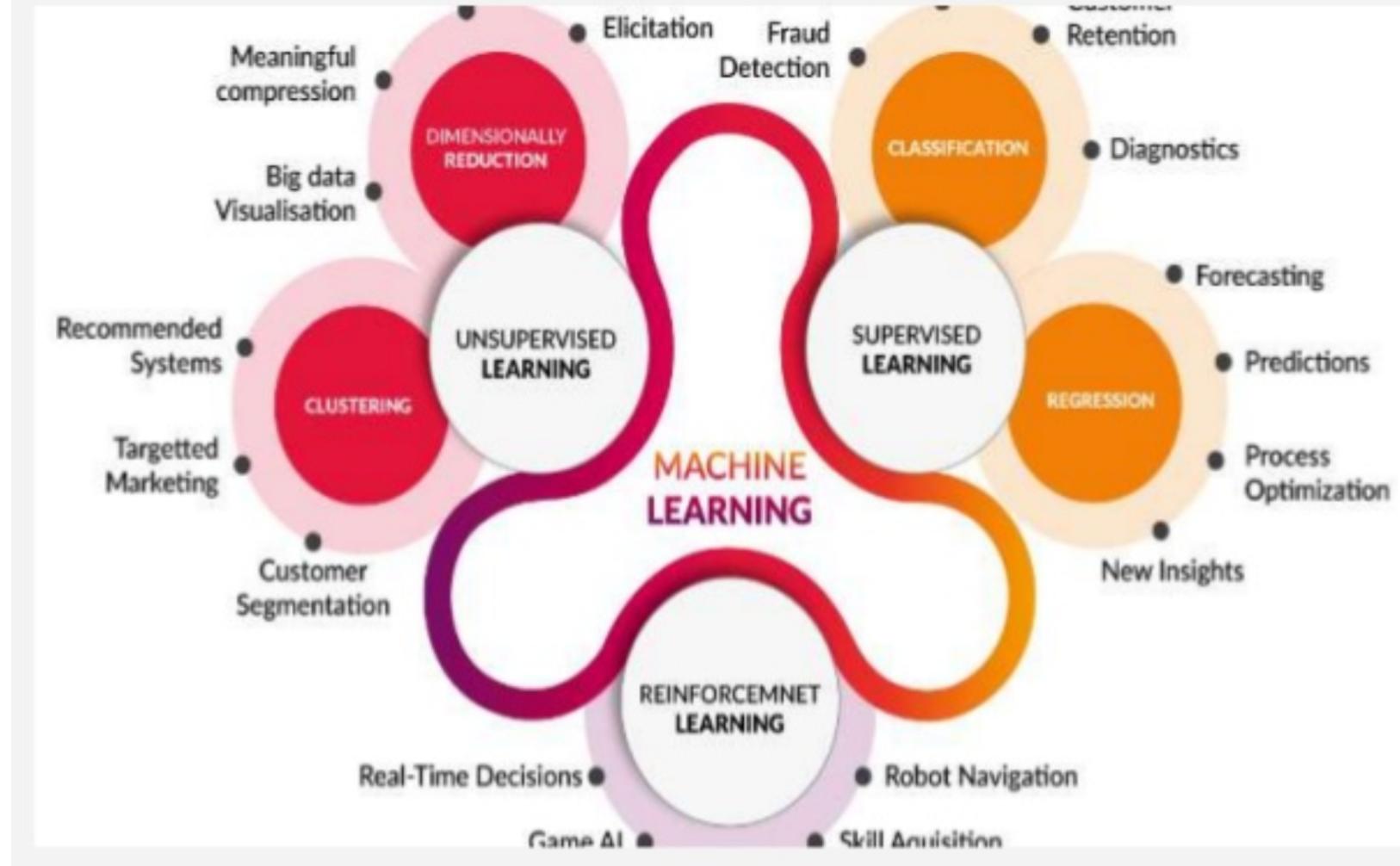
- Statistics problems
- Numerical problems
- Open source libraries
- Statistics, Linear algebra, Standard deviation, Slope

Python libraries

- NumPy - <https://numpy.org/>
- Pandas - <https://pandas.pydata.org/>
- Matplotlib - <https://matplotlib.org/>
- Seaborn - <https://seaborn.pydata.org/>
- Sklearn - <https://scikit-learn.org/>

and many more...

Various algorithms and selection of algorithms



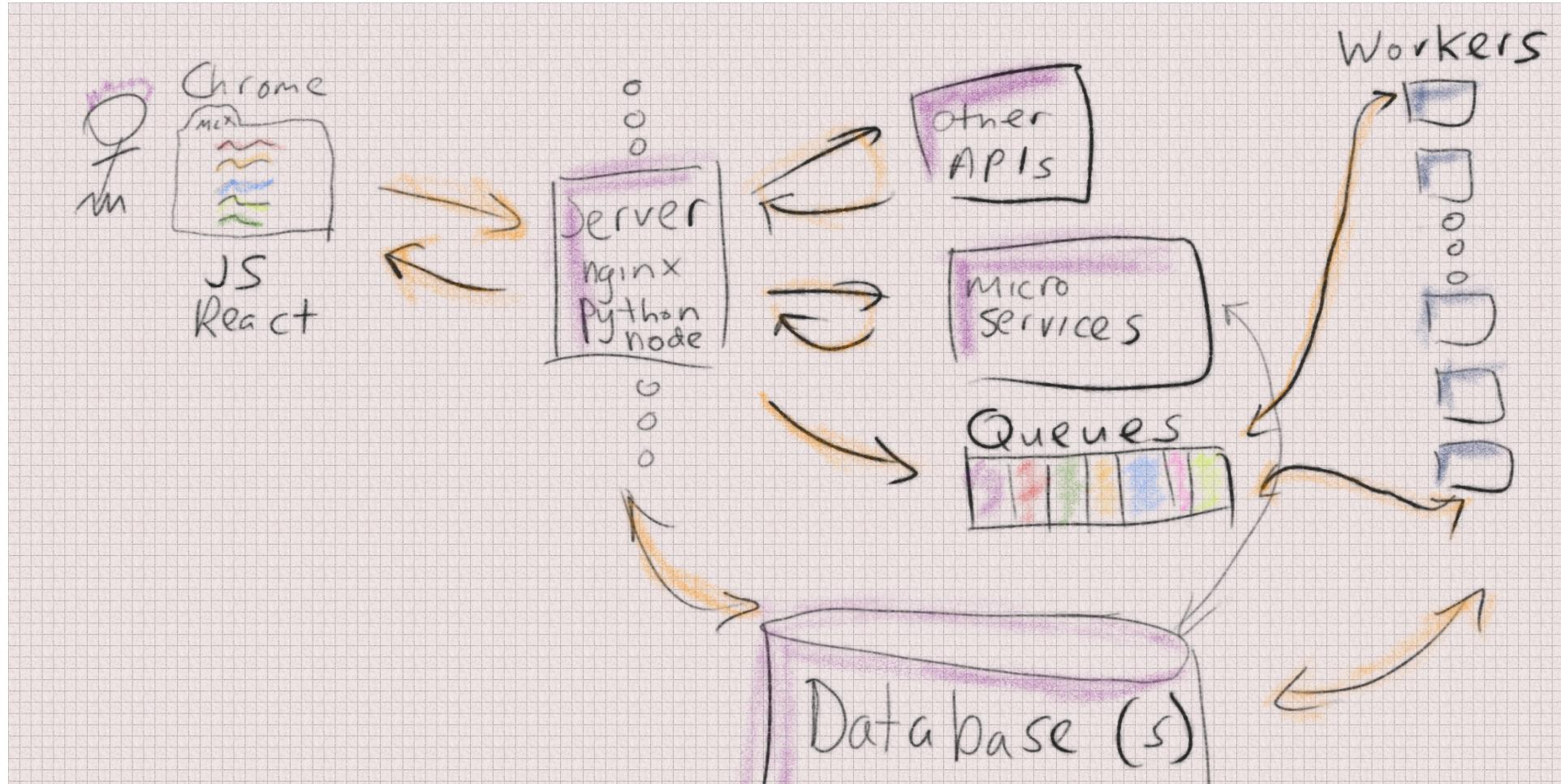
Model evaluation and boosting

- Improve by More Data (Rows)
- Cross Validation (Train Test Split)
- Improve by filling missing value, removing outlier, pre-processing
 - Transform
 - Standardize or Normalize
- Improve accuracy by Choosing better Algorithm
- Improve accuracy by hyperparameters of Algorithms
- Improve Accuracy by Combine different models
 - Bagging
 - Boosting
 - Voting
 - Stacking

Improve model training, Feature engineering

- Improve Training time
 - Dimensionality Reduction
 - Multicollinearity
 - Constant (Low Variance features)
 - PCA, LDA
 - Improve by Feature Engineering (Columns)
 - Create new features
 - Removing irrelevant features
 - Feature transform
 - Features from Dates
 - Features from Numerical variables - Averages etc
 - Interaction variables

Deployment and Maintenance



Myths of ML and AI

- AI and machine learning are the same thing
- More training data leads to better AI outcomes
- AI can work effectively on any or all the messy data
- AI and machine learning are basically just software development
- AI replaces the need for human intelligence
- AI works like the human brain
- Intelligent machines learn on their own
- AI and ML are latest technologies

Points to be remembered while doing ML

- Encryption
- Privacy use data
- Reproducibility
- Adversarial attacks on the model.
- Impact of predicting patterns.

Google, Twitter, Hike, Amazon use cases

- Google Maps, Google Photos, Diabetic retinopathy
- Tweet suggestions & priorities, Hike stickers & Natas
- Amazon/Flipkart product suggestions
- Photo apps, AI camera
- Reebok, Adidas smart showroom
- Loreal Try on – ModiFace model
- Lenskart – SunGlass model
- Harmonies in piano, Shadow art, Text and speech recognition

Current research areas

- Make solar energy affordable
- Engineer better medicines
- Develop carbon sequestration methods
- Manage Nitrogen cycle
- Reverse engineer brain
- Secure cyberspace
- Enhance virtual reality
- Advance health informatics

Current research areas

- Provide access to clean water
- Provide energy from fusion
- Restore and improve urban infrastructure
- Prevent nuclear terror
- Advance personalized learning
- Engineer the tool for scientific discovery

Build for digital India program

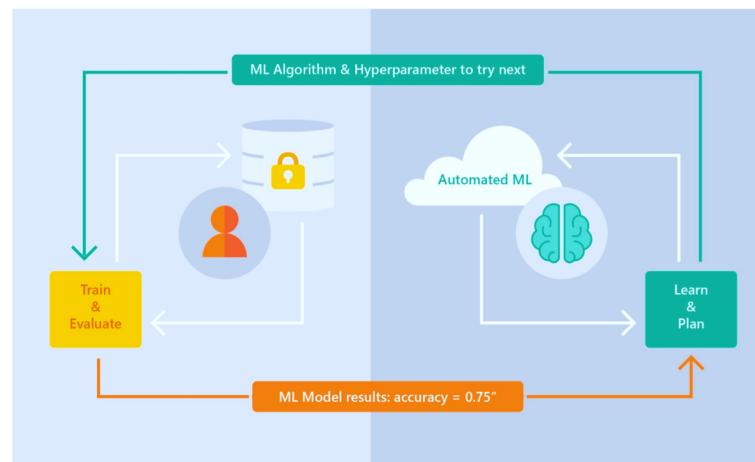
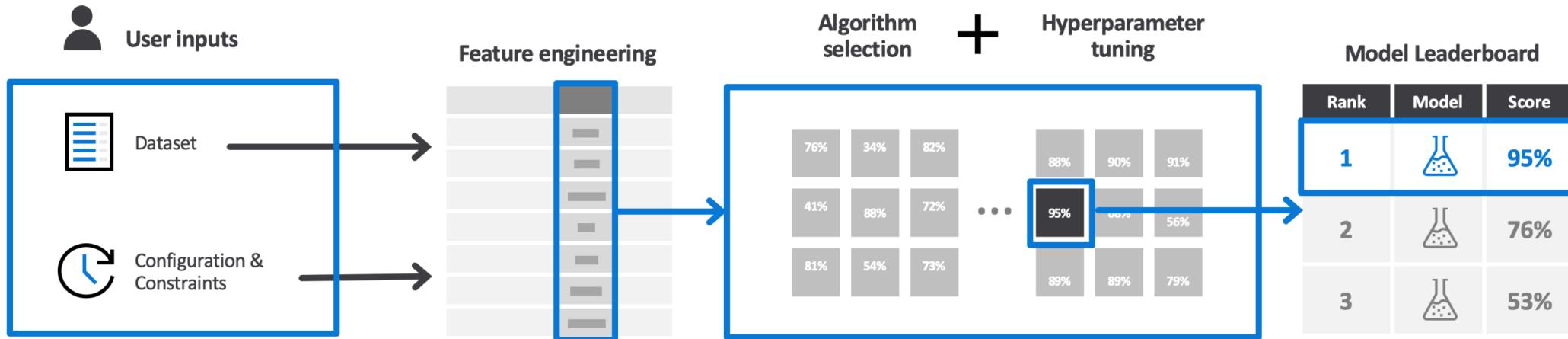
- <http://bit.ly/buildfordigitalindia>
- [AI research lab in India](#)

ML dev leads summit - India

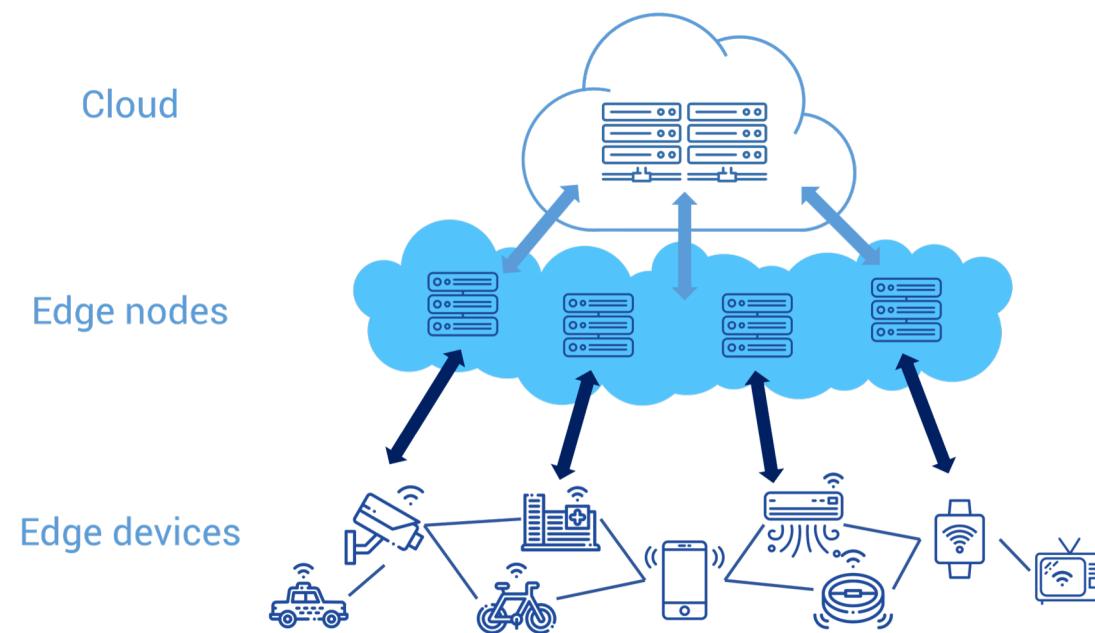
- ML with Industries, Students, Society
- ML frameworks analysis
- Road map plans & Open source contribution
- Research work collaboration & testing
- Community meetups, Hackathons, Study jams
- Debugging and Testing ML models
- Design pattern for neural network
- Startups support



Auto ML and ML.NET



CPU, TPU, GPU and Edge computing



Jupyter notebook

- <https://jupyter.org/>



Google colab

- <https://colab.research.google.com/>
- <https://www.tensorflow.org/tutorials/keras/classification>

The Colab logo consists of the word "colab" in a bold, sans-serif font. The letters are primarily orange, with the "c" and "o" partially overlaid by a yellow semi-circle at the bottom.

Free TPUs for ML research

- <https://g.co/tputalk>
- 5 regular Cloud TPUs + 20 preemptible Cloud TPUs per person
- 1 Cloud TPU ≈ 4 v100 GPUs
- Free access available for several months
- <https://goo.gl/Amd9qQ> - To say ‘Hello to TPU’
- <https://goo.gl/dnW4wp> - To try in on **Keras**

TensorFlow libraries

- <https://www.tensorflow.org/resources/libraries-extensions>



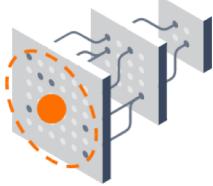
Sample Machine learning problem in Colab

- https://colab.research.google.com/github/google/eng-edu/blob/master/ml/pc/exercises/image_classification_part1.ipynb



What is new in TensorFlow 2.0

- * Tf.text and NLP
- * TFLite for edge devices
- * TensorFlow.js
- * Swift for TensorFlow
- * TensorFlow Extended
- * Model optimization and Quantization



Easy model building

Build and train ML models easily using intuitive high-level APIs like Keras with eager execution, which makes for immediate model iteration and easy debugging.



Robust ML production anywhere

Easily train and deploy models in the cloud, on-prem, in the browser, or on-device no matter what language you use.



Powerful experimentation for research

A simple and flexible architecture to take new ideas from concept to code, to state-of-the-art models, and to publication faster.

TensorFlow Lite

- <https://www.tensorflow.org/lite/>



Pick a model

Pick a new model or retrain an existing one.



Convert

Convert a TensorFlow model into a compressed flat buffer with the TensorFlow Lite Converter.



Deploy

Take the compressed .tflite file and load it into a mobile or embedded device.

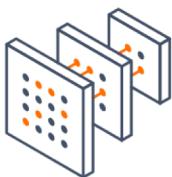


Optimize

Quantize by converting 32-bit floats to more efficient 8-bit integers or run on GPU.

TensorFlow for JS (Demo)

- <https://www.tensorflow.org/js/>



Run existing models

Use off-the-shelf JavaScript models or convert Python TensorFlow models to run in the browser or under Node.js.



Retrain existing models

Retrain pre-existing ML models using your own data.



Develop ML with JavaScript

Build and train models directly in JavaScript using flexible and intuitive APIs.

TensorFlow Extended



TensorFlow Data Validation

[Get started →](#)

TensorFlow Data Validation (TFDV) helps developers understand, validate, and monitor their ML data at scale. TFDV is used to analyze and validate petabytes of data at Google every day, and has a proven track record in helping TFX users maintain the health of their ML pipelines.



TensorFlow Transform

[Get started →](#)

When applying machine learning to real world datasets, a lot of effort is required to preprocess data into a suitable format. This includes converting between formats, tokenizing and stemming text and forming vocabularies, and performing a variety of numerical operations such as normalization. You can do it all with `tf.Transform`.



TensorFlow Model Analysis

[Get started →](#)

TensorFlow Model Analysis (TFMA) enables developers to compute and visualize evaluation metrics for their models. Before deploying any machine learning (ML) model, ML developers need to evaluate model performance to ensure that it meets specific quality thresholds and behaves as expected for all relevant slices of data. For example, a model may have an acceptable AUC over the entire eval dataset, but underperform on specific slices. TFMA gives developers the tools to create a deep understanding of their model performance.



TensorFlow Serving

[Get started →](#)

Machine Learning (ML) serving systems need to support model versioning (for model updates with a rollback option) and multiple models (for experimentation via A/B testing), while ensuring that concurrent models achieve high throughput on hardware accelerators (GPUs and TPUs) with low latency. TensorFlow Serving has proven performance handling tens of millions of inferences per second at Google.

TensorFlow Extended - workshop

- <https://goo.gle/tfx-workshop-slides>

Where to start ? How to start ?

Books / E-learning

- Basic
 - kaggle.com/learn
 - <https://www.dataschool.io/machine-learning-with-scikit-learn/>
 - <https://www.udemy.com/python-for-data-science-and-machine-learning-bootcamp>
 - A Concise Handbook of TensorFlow 2.0
- Intermediate
 - “Data Science for Executives coursera” & “AI for everyone coursera”
 - <https://www.coursera.org/specializations/machine-learning>
 - <https://www.coursera.org/specializations/data-science-python>
 - “How to Win a Data Science Competition: Learn from Top Kagglers coursera”
 - Book: Hands-On Machine Learning with Scikit Learn and TensorFlow O'Reilly

Journals/papers

- Important Papers
 - <https://adesshpande3.github.io/adesshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>
 - <http://jmlr.org/papers/v15/delgado14a.html>
- Keeping updated with Latest Research
 - <https://arxiv.org/>
 - <http://www.arxiv-sanity.com/> (**Very Useful resource. Automatically recommends new papers according to your interests**)
 - <http://www.gitxiv.com/> (Source code of arxiv papers)
- Summary of good papers
 - <https://github.com/dennybritz/deeplearning-papernotes>
 - <https://medium.com/paper-club>
 - <https://www.youtube.com/channel/UCNIkB2IeJ-6AmZv7bQ1oBYg> (Arxiv Insights)
 - <https://www.youtube.com/user/keeroyz> (2 minutes papers)

Blogs, documentation, Datasets

- Machine Learning Intermediate
 - <http://course.fast.ai/ml.html>
 - <https://towardsdatascience.com/>
- Machine Learning using TensorFlow
 - <https://developers.google.com/machine-learning/crash-course/>
 - Coursera: Machine Learning with TensorFlow on Google Cloud Platform Specialization
- Medium, Hackernoon blogs, GitHub documentation
- Datasets
 - <http://academictorrents.com/>
 - <https://github.com/awesomedata/awesome-public-datasets>
 - <https://gengo.ai/articles/the-50-best-free-datasets-for-machine-learning/>
 - <https://www.forbes.com/sites/bernardmarr/2018/02/26/big-data-and-ai-30-amazing-and-free-public-data-sources-for-2018>

ML career path

- Full stack Machine Learning Engineer
 - Data Engineer
 - Machine learning evangelist
 - Machine Learning developer
 - Machine Learning consultant
 - Domain consultant
 - Exploratory data analyst
- etc..etc..etc..

Acknowledgements



- **Andrew Ng** - @AndrewYNg
- **Laurence Moroney** - @lmoroney
- **Ajinkya Kolhe** - @ajinkyakolhe112
- **karthikeyan VK** - @karthik3030
- **Google Devs India** - @GoogleDevsIN
- **Microsoft Developer** - @msdev
- **TensorFlow** - @TensorFlow
- **Infosys Nia** - @PurposefulAI and many more...!

It's Question time...!

Feedback...!

Thank you...!