Part 1

MACHINE LEARNING SOLUTIONS FOR BUSINESS PROBLEMS

Basic Machine Learning Functionality
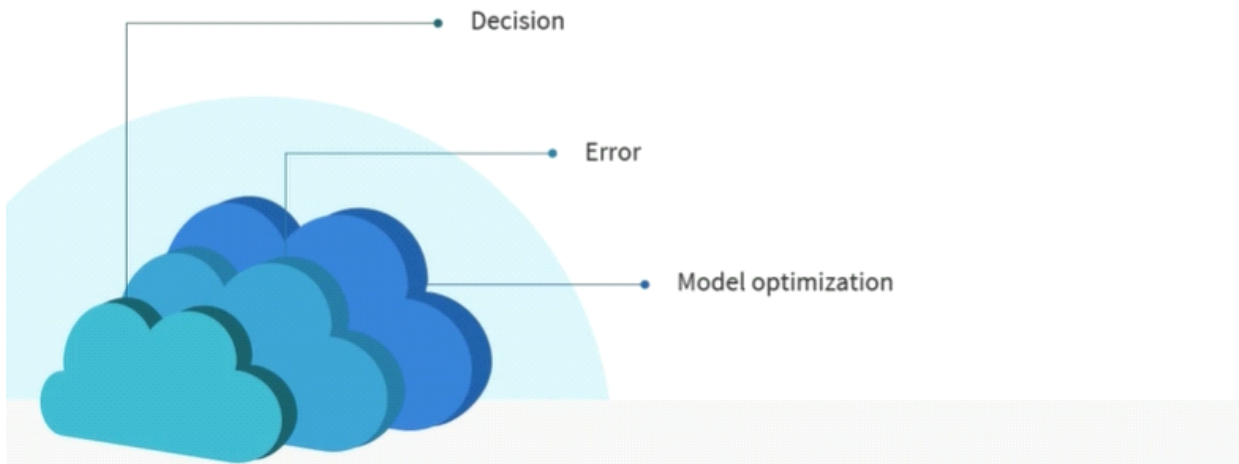
DIFFERENCE IN MACHINE LEARNING AND AI
- Machine learning is a subset of AI
  - End goal is to analyses large models of data
- Methodology
  - ML
    - Supervised
      - Uses labeled data as input
    - Unsupervised
      - Label with no input
      - ML model tries to asses what data is being used,

  - AI uses neural networks (genetic algorithms, deep learning, search based learning)
- Implementation
  - Find training set by cleaning the dataset remove outliers
  - Choose preexisting strategy or model
    - Linear regression or decision tree
  - Almost off the shelf solution
- Requirements
  - Require several hundred data points
  - You can use single server or cluster of servers
  - Computationally expensive

Machine learning solutions for business
- Customer segmentation based on behavior for example,
- Fraud detection to identify when someone is committing a fraudulent act
- Sentiment analysis
  - Looking at the text from customer feedback and analyze the customer

SUPERVISED LEARNING

Requires supervision

Trained on labeled data

Key applications
- Data classification
- Trend prediction/ forecasting
    ○ Regression
    ○ Data classification
- Change forecasting

UNSUPERVISED MACHINE LEARNING

Trained on unlabeled data

- We have to clean it to do statistical analysis

Key applications
- Segmenting data
- Customer trend identification
- Automated recommendations
- Personalized content
- Systems and process efficiency

SEMI SUPERVISED MACHINE LEARNING
- Combines elements from both SL and UNSL
- Trained on subset of correctly labelled data and then use the model to label the unlabeled data
- It can be used to understand the relationship between data points

Key applications
- Categorizing data
- Processing media files (audio or image files)
- Group large amounts of text

REINFORCEMENT MACHINE LEARNING
- Uses trial and error
- Successful actions are rewarded

Key applications
  - AI development
  - NLP
  - Self-driving vehicles
  - Video game AI
  - Finding efficiencies
  - Playing games against humans

EXAMPLES OF ML ALGORITHMS IN ACTION
  - Neural networks
  - Linear Regression
  - Logistic Regression
    ○ SL used for categorical responses (0/1, T/F, YES/NO)
  - Decision trees
    ○ Predict numerical values
    ○ Classify data into categories
  - Random Forests
    ○ If you have multiple decision trees you can group them together
  - Clustering
    ○ Help identify differences between the data
    ○ Two classes that are physically different then if you can put them on a graph and there
      will be clusters formed and can help segment the data

Examples of Machine Learning Algorithms in Action

- Neural networks
- Linear regression
- Logistic regression
- Decision trees
- Random forests
- Clustering
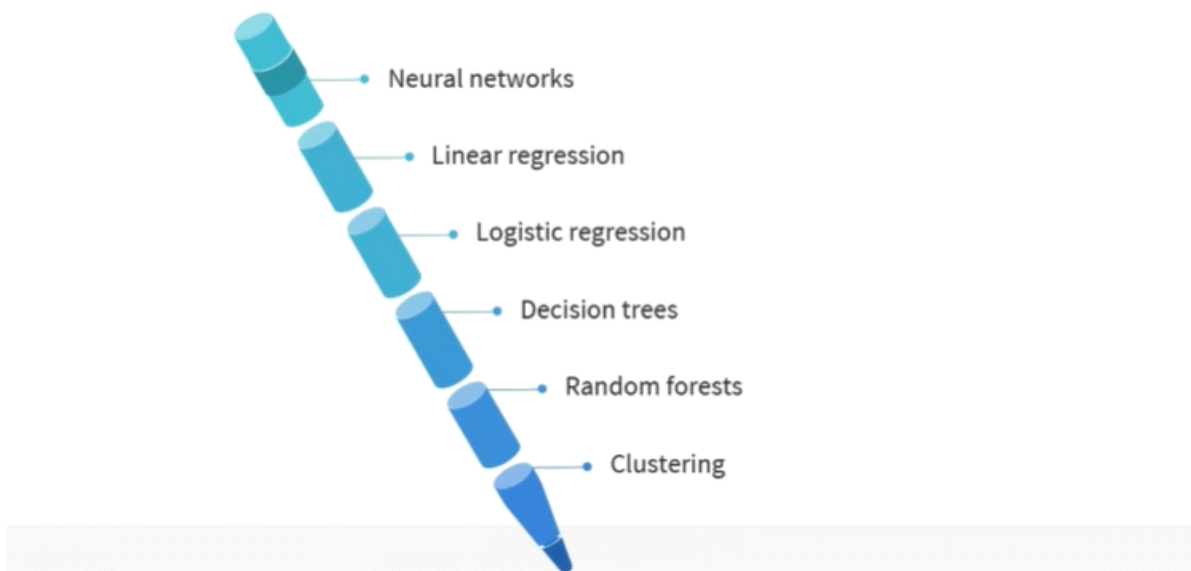
BASICS OF ML Workflow

1) **Data Gathering**
2) **Data Pre-processing**
3) **Data construction**
   a. **Feature validation / feature engineering**
4) **Training and Refinement**
5) **Evaluation step**
   a. Measure and test the model accuracy

We take this workflow and talk about it in Vertex AI model training
  - AutoML
  - Custom Training
  - Ray on Vertex AI

Vertex AI Model Training

-

AutoML

Custom training

Ray on Vertex AI

which lets us use Ray to sca
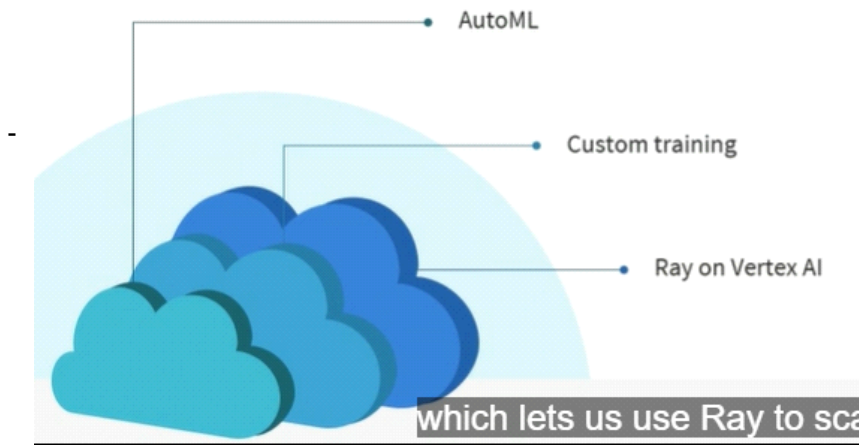
<u>Another - Big Query Enterprise Data Warehouse</u>
- Is a fully managed enterprise tool, data warehouse tool
- Can analyze the data directly in that tool
- All powered by SQL and data scientists primarily use SQL
- They can run large queries
- All serverless
- And contain built in machine learning models
- Can help batch predictions on models

Theres an automation step in Vertex AI and ML

1) Automate
    a. Automate the ingestion step
2) After the model Is trained deploy it and monitor the model
3) Manage the model itself and take it down or bring it up
    a. The infrastructure of model can be managed directly in the vertex pipeline

Machine learning pipeline Life cycle
- Define the ML pipeline lifecycle
- Compile the code to run directly in the system
    ○ Dockerize the code and send it to pipeline to run training loop
- Run the code in pipeline
    ○ Take the docker image and open it up and put into cloud
- After creating pipeline run we can
    ○ Monitor the system performance
    ○ Visualize (graphs, metrics)
    ○ Analyze our datasets
   ○ Stop or delete pipeline runs

Machine Learning Pipeline Life Cycle

Monitor, visualize, analyze

Stop or delete pipeline runs

AI and ML products

1) Vertex AI studio
2) Vertex AI agent Builder
   a. This builder provdes low code for orchestration and customization capabilities
   b. Can be used to build a good search experience on my own data
   c. You can use google vertex ai builder
   d. Theres generative AI document summarization
      i. It creates summary using vertex generative AI studio
         1) Stores the searchable summary in BigQuery database so that you can query
            It with applications when necessary

ML and ML operations tools
- Used to create, test and monitor, tune and deploy machine learning and AI models
- It has built into
  o vertex AI platforms,notebooks
    ▪ Can create custom ML training loops
    ▪ :) the notebooks are shareable
    ▪ :) rapid protoyping and model development with little overhead
  o AutoML

  o

Vertex AI platform	Vertex AI Notebooks	AutoML
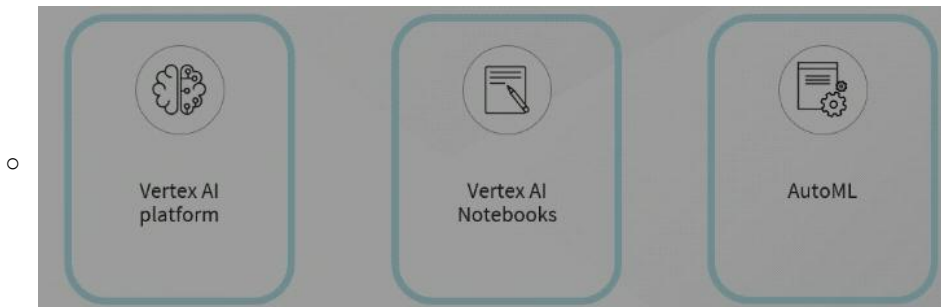
    ▪ Allows to build custom ML with minimal effort
    ▪ Custom machine learning models in minutes with little code
    ▪ Can train models specific to business needs
    ▪ AutoML will figure out the rest and the best way to train it
- **Speech , Text, Language APIs**

  Natural language application programming interface (API)

- **NLP API**
  o **Can be applied** to understand apps with the api

- o Train an open machine learning models to classifcy, extract and detect sentiment from language
- o Text to speech is powerful

o

- o Can also be used for speech to text, where you talk into a microphone and write down in text (closed caption thing) so accurately convert speech into text using API
- o Great for automatic speech recognition and real time transcription

o

- Translation AI
    - o We can use this content in apps for multilingual, fast, dynamic machine translation
    - o Great for real time translation
    - o Great for localisation of content so we can internationalize our products much easier using this artificial intelligence.

IMAGE AND VIDEO APIs

- Vision AI
    - o We can derive insights from our images in the cloud with auto machine learning vision, or use pretrained vision API models to detect objects, understand text more accurately.
- Video AI
    - o We also have video AI which is enabled powerful content discovery and engaging video

experiences
- ○ We can extract rich metadata at the video shot or frame level
- ○ This is all done through custom AutoML tools.

Video AI is very useful for analyzing videos and vision AI is good for high resolution pictures

Document and Data APIs
- So that we can search and store our documents
- Easy to for us to extract classify and split our Data from our documents
- Reducing manual document processing
- Makes things a little bit easier for us when we gain insights from our document data this way

AI Assistance and Conversational AI
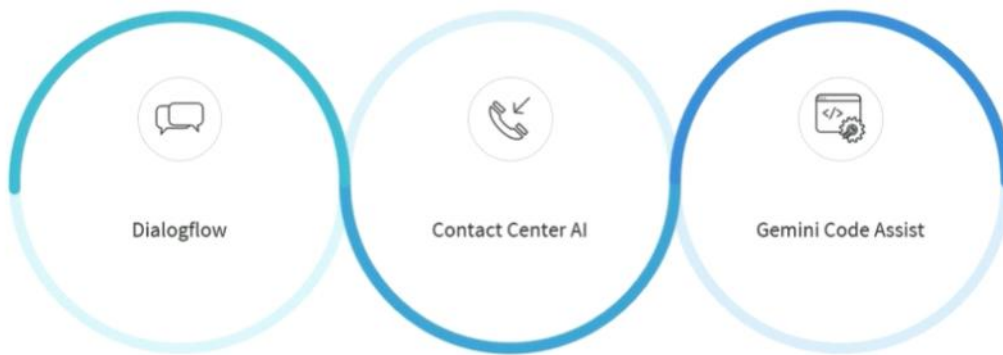
- Dialog flow
    - ○ Conversational AI platforms great for AI and LLM
    - ○ This builds natural interactions for complex multiturn conversations builds and deploys advanced agents quickly
    - ○ These are great for chatbots
    - ○ They are very advanced
    - ○ We can build and manage these virtual agents with dialog flow
    - ○ Each agent can be an expert or chatbot on one topic
- Contact Center AI
    - ○ We can transform our contact center with AI technology, assisted with Dialog Flow + agent assist
    - ○ So we can increase operational efficiency and personalized customer care with this tool
    - ○ Its end to end customer care as a service solution with its own call center solution.
- Gemini code assist
    - ○ It offers code recommendations in real time
    - ○ Suggest full functionality and code blocks
    - ○ Identifies vulnerabilities in our code as we write it
    - ○ Suggest fixes as its noticing while write
    - ○ Keeps code cleaner
    - ○ and helps find errors faster
    - ○ Allows me to identify errors much faster
    - ○ Available for Go, Java, JavaScript, Python, SQL
    - ○ You can have extensions in your IDE editor - VSCode,JetBrains

AI infrastructure

- **Tensor processor units (TPUs), Graphics processing units (GPUs), central processing units (CPUs)**
    - ○ TPU - Work with neural networks to train them
    - ○ GPU - great for video analysis and image analysis
    - ○ CPU - great for ML and mathematics

- **Google Kubernetes Engine (GKE)**
  - Allows us to scale the images
  - With more demand and images coming in we want to deploy to that area where it will meet demand
  - If there's a lot of people trying to use our model to retrieve an inference
    - Submitting text , image to a model
    - It will get spin up extra machines as traffic is high and route the traffic to those machines and so that means we will have horizontally scaled solutions
    - When there's no traffic there's no need to pay for it
    - It will slow things down by taking away the machines that don't need to be there
    - That's a way for it to intelligently make sure we are not overwhelming our machine
    - It allows us to save costs and horizontally scale the machine based on our demand.