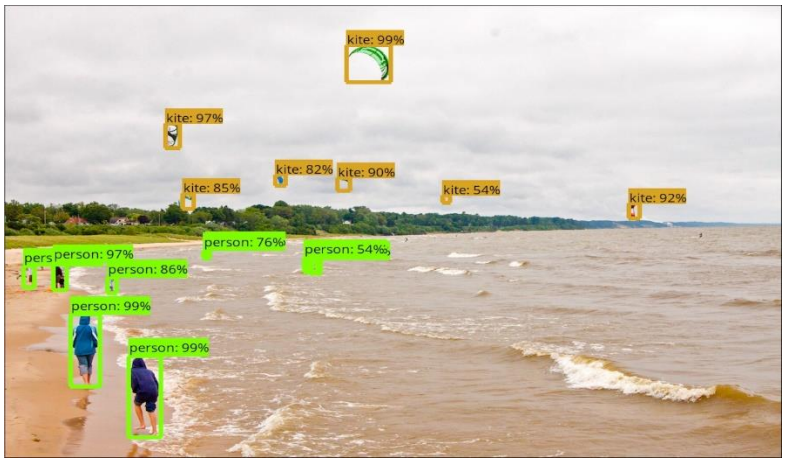
Question #1 - What is TensorFlow Object Detection API?

Answer #1 - TensorFlow Object Detection API is designed to identify and localize multiple objects within an image.



For any further detail:

https://github.com/tensorflow/models/tree/master/research/object_detection

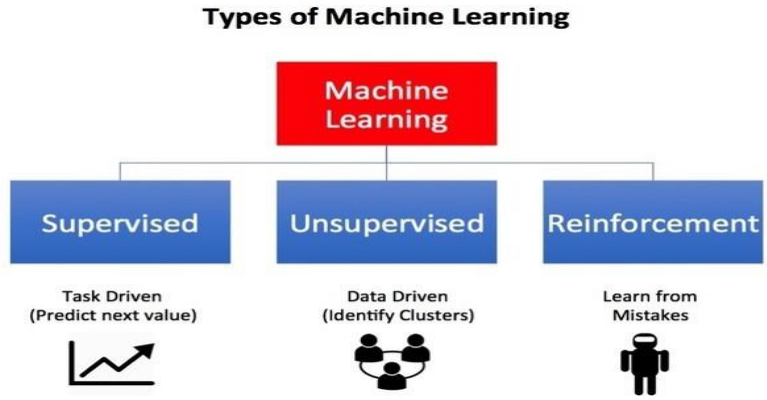
https://cloud.google.com/ai-platform/training/docs/algorithms

https://cloud.google.com/ai-platform/training/docs

https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/img/kites_detections_output.jpg

Question #2 – What is Q-learning is an RL Reinforcement Learning algorithm?

Answer #2 - Q-learning is an RL Reinforcement Learning algorithm. RL provides a software agent that evaluates possible solutions through a progressive reward in repeated attempts. It does not need to provide labels. But it requires a lot of data and several trials and the possibility to evaluate the validity of each attempt. The main RL algorithms are deep Q-network (DQN) and deep deterministic policy gradient (DDPG).

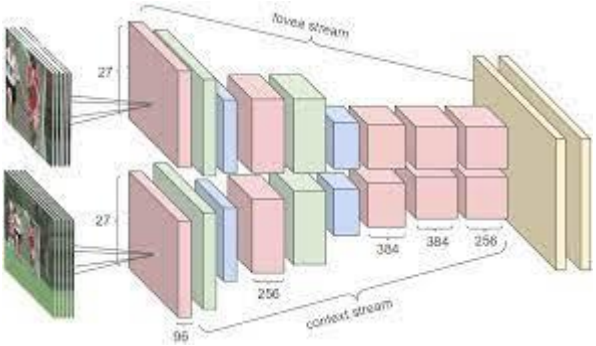


https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292

Question #3 – What is Convolutional Neural Network?

Answer #3 - A Convolutional Neural Network is a Deep Neural Network in which the layers are made up of processed sections of the source image. This technique allows you to simplify images and highlight shapes and features regardless of the physical position in which they may be found.

For example, if we have the same signature in the center or at the bottom right of an image, the object will be different. But the signature is the same. A neural network that compares these derived features and can simplify the model achieves the best results.



For any further detail:

Convolutional Neural Networks — A Beginner's Guide | by Krut Patel

https://research.google.com/pubs/archive/42455.pdf

Question #4 – What is Feature Cross?

Answer #4 - A cross of functions is a dome that creates new functions by multiplying (crossing) two or more functions. It has proved to be an important technique and is also used to introduce non-linearity to the model.

Question #5 – What Bigquery ML ARIMA does?

Answer #5 - Bigquery ML ARIMA_PLUS can manage time-series forecasts. The model automatically handles anomalies, seasonality, and holidays.

Question #6 – What Dataflow Flex does?

Answer #6 - Dataflow Flex to aggregate and extract insights in real-time in BigQuery

For any further detail:

https://cloud.google.com/solutions/building-anomaly-detection-dataflow-bigqueryml-dlp

https://cloud.google.com/architecture/detecting-anomalies-in-financial-transactions

Question #7 – What is K-Means algorithm?

Answer #7 - The k-means clustering is a mathematical and statistical method on numerical vectors that divides ann observes k clusters. Each example belongs to the cluster with the closest mean (cluster centroid).

In ML, it is an unsupervised classification method and is widely used to detect unusual or outlier movements. For these reasons, it is one of the main methods for fraud detection.

But it is not the only method because not all frauds are linked to strange movements. There may be other factors.

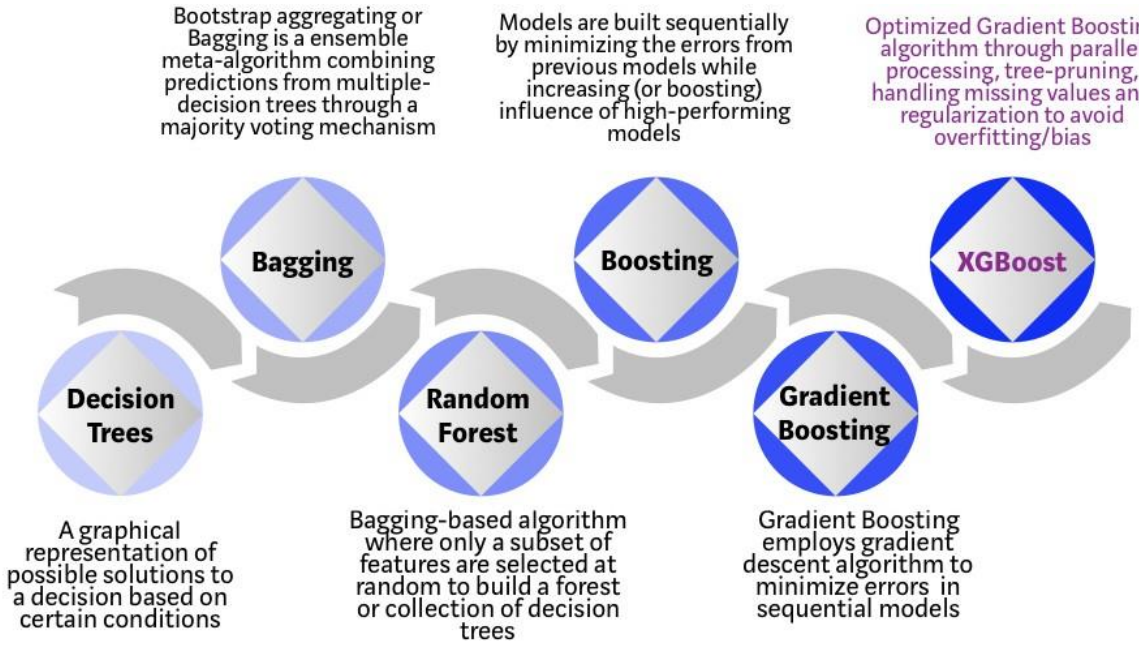
For any further detail:

https://cloud.google.com/solutions/building-anomaly-detection-dataflow-bigqueryml-dlp

https://cloud.google.com/architecture/detecting-anomalies-in-financial-transactions

Question #8 – What is Boosted Tree – XGBoost algorithm?

Answer #8 - XGBoost, which as you can see from the figure, is an evolution of the decision trees, has recently been widely used in this field and has had many positive results.



It is an open-source project and this is the description from its Github page:

XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solves many data science problems in a fast and accurate way. The same code runs on major distributed environments (Kubernetes, Hadoop, SGE, MPI, Dask) and can solve problems beyond billions of examples.

For any further detail:

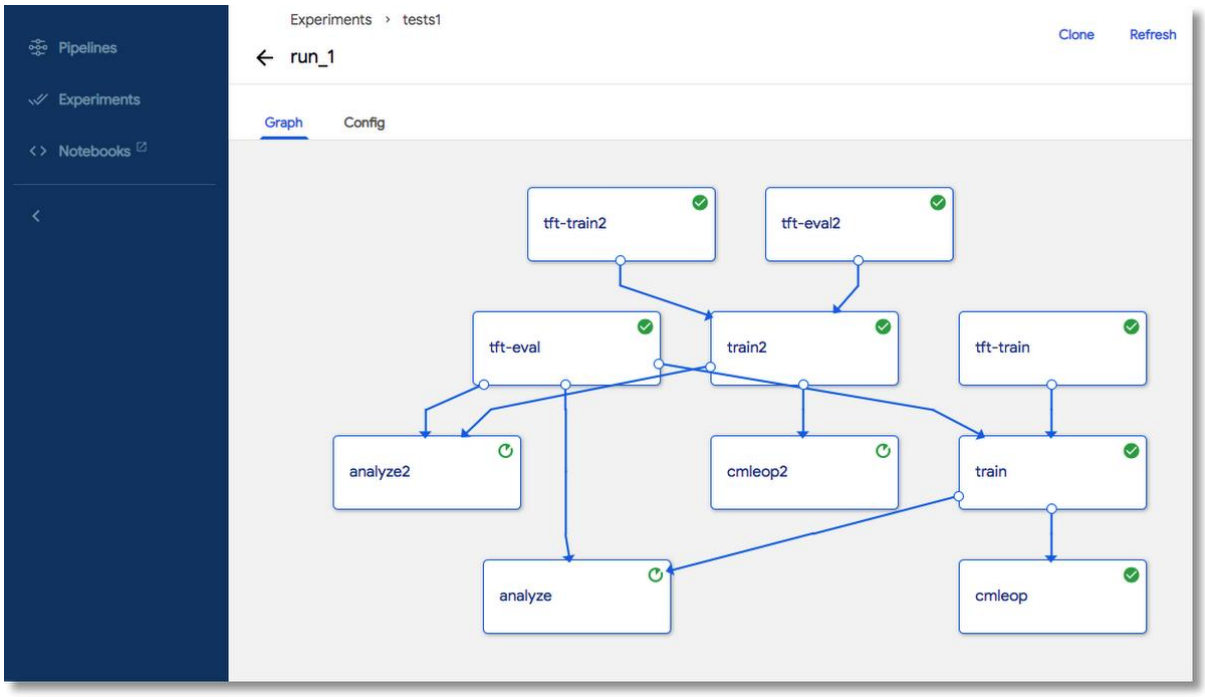
https://medium.com/@adityakumar24jun/xgboost-algorithm-the-new-king-c4a64ea677bf

Question #9 – What is Kubeflow?

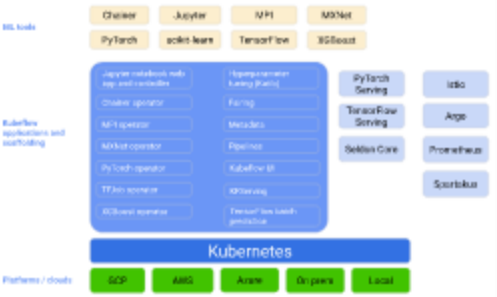
Answer #9 - Kubeflow Pipelines is an open-source platform designed specifically for creating and deploying ML workflows based on Docker containers.

Their main features:

- Using packaged templates in Docker images in a K8s environment
- Manage your various tests/experiments
- Simplifying the orchestration of ML pipelines
- Reuse components and pipelines



It is within the Kubeflow ecosystem, which is the machine learning toolkit for Kubernetes.



For any further detail:

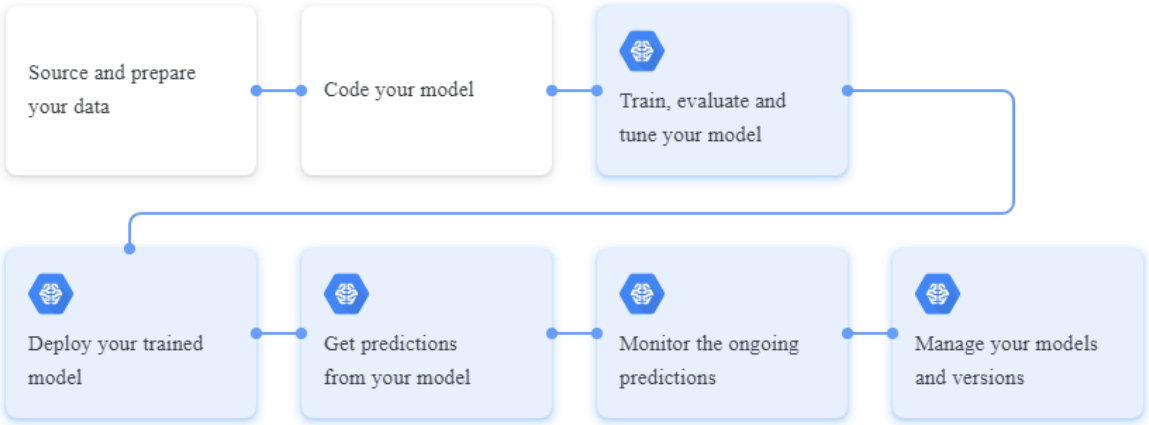
https://cloud.google.com/ai-platform/training/docs/tensorflow-2

https://cloud.google.com/ai-platform/docs/technical-overview

Question #10 – What is Vertex AI?

Answer #10 – Vertex AI is an integrated suite of ML-managed products aimed at:

- Train an ML model
- Evaluate and tune a model
- Deploy models
- Manage prediction: Batch, Online and monitoring
- Manage model versions: workflows and retraining



For any further detail:

https://cloud.google.com/ai-platform/training/docs/tensorflow-2

https://cloud.google.com/ai-platform/docs/technical-overview

Question #11 – What is Cloud Data Loss Prevention API (DLP)?

Answer #11 - Cloud Data Loss Prevention is a managed service specially designed to discover sensitive data automatically that may be protected. It could be used for personal codes, credit card numbers, addresses, and any private contact details, etc.

For any further detail:

https://cloud.google.com/architecture/sensitive-data-and-ml-datasets

Question #12 – What is Cloud Speech API?

Answer #12 - Cloud Speech API is useful if you have audio recordings as it is a speech-to-text service.

Question #13 – What is Cloud Vision API?

Answer #13 - Cloud Vision API has a built-in text-detection service. So, you can get text from images.

Question #14 – What to do if you have imbalance data in classification problem?

Answer #14 - When there is an imbalance between true and false ratios in binary classification, it is necessary to modify the classification threshold so that the most probable errors are those with minor consequences. In our case, it is better to be wrong with a healthy person than with a sick one.

Accuracy is the number of correct predictions on the total of predictions done.

Let's imagine that we have 100 predictions, and 95 of them are correct. That is 95%. It looks almost perfect.

But we assume that the system has foreseen 94 true negative cases and only one true positive case, and one case of false positive, and 4 cases of false negative.

So, the model predicted 98 healthy when they were 95 and 2 suspected cases when they were 5.

The problem is that sick patients are, luckily, a minimal percentage. But it is important that they are intercepted. So, our model failed because it correctly identified only 1 case out of the total of 5 real positives that is 20% (recall). It also identified 2 positives, one of which was negative, i.e. 50% (precision).

It's not good at all.

Precision: Rate of correct positive identifications

Recall: Rate of real positives correctly identified

To calibrate the result, we need to change the threshold we use to decide between positive and negative. The model does not return 0 and 1 but a value between 0 and 1 (sigmoid activation function). In our case, we have to choose a threshold lower than 0.5 to classify it as positive. In this way, we risk carrying out further investigations on the healthy but being able to treat more sick patients. It is definitely the desired result.

For any further detail:

https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

Question #15 – What is Document AI?

Answer #15 - Document AI is the ideal broad-spectrum solution. It is a service that gives a complete solution with computer vision and OCR, NLP and data management. It allows you to extract and structure information automatically from documents. It can also enrich them with the Google Knowledge Graph to verify company names, addresses, and telephone numbers to draw additional or updated information.

https://cloud.google.com/document-ai

Question #15 – What normalization techniques could you use?

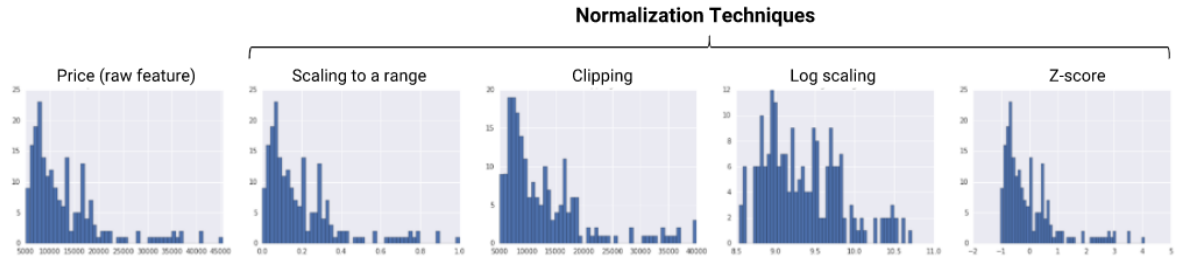
Answer #15 - **Feature clipping** eliminates outliers that are too high or too low.

Scaling means transforming feature values into a standard range, from 0 and 1 or sometimes -1 to +1. It's okay when you have an even distribution between minimum and maximum.

When you don't have a fairly uniform distribution, you can instead use **Log Scaling** which can compress the data range: x1 = log (x)

Z-Score is similar to scaling, but uses the deviation from the mean divided by the standard deviation, which is the classic index of variability. So, it gives how many standard deviations each value is away from the mean.

All these methods maintain the differences between values, but limit the range. So the computation is lighter.



For any further detail:

https://developers.google.com/machine-learning/data-prep/transform/normalization

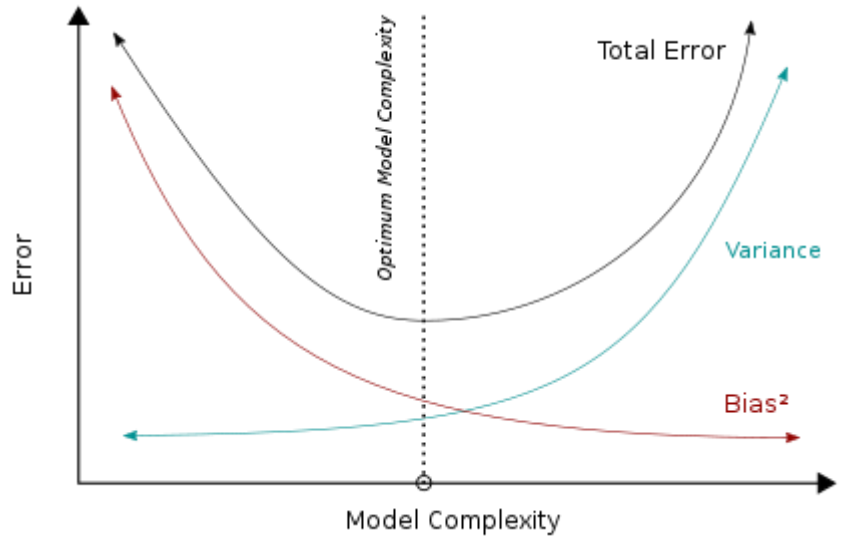
Question #17 – What is Bias-Variance tradeoff?

Answer #17 - The variance indicates how much function f (X) can change with a different training dataset. Obviously, different estimates will correspond to different training datasets, but a good model should reduce this gap to a minimum.

The bias-variance dilemma is an attempt to minimize both bias and variance.

The bias error is the non-estimable part of the learning algorithm. The higher it is, the more underfitting there is.

Variance is the sensitivity to differences in the training set. The higher it is, the more overfitting there is.



For any further detail:

https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff

Question #18 – What is BigQueryML?

Answer #18 - BigQuery and BigQueryML are powerful services for data analysis and machine learning. They are fully serverless services that can process petabytes of data in public and private datasets and even data stored in files.

BigQuery works with standard SQL and has a CLI interface: bq.

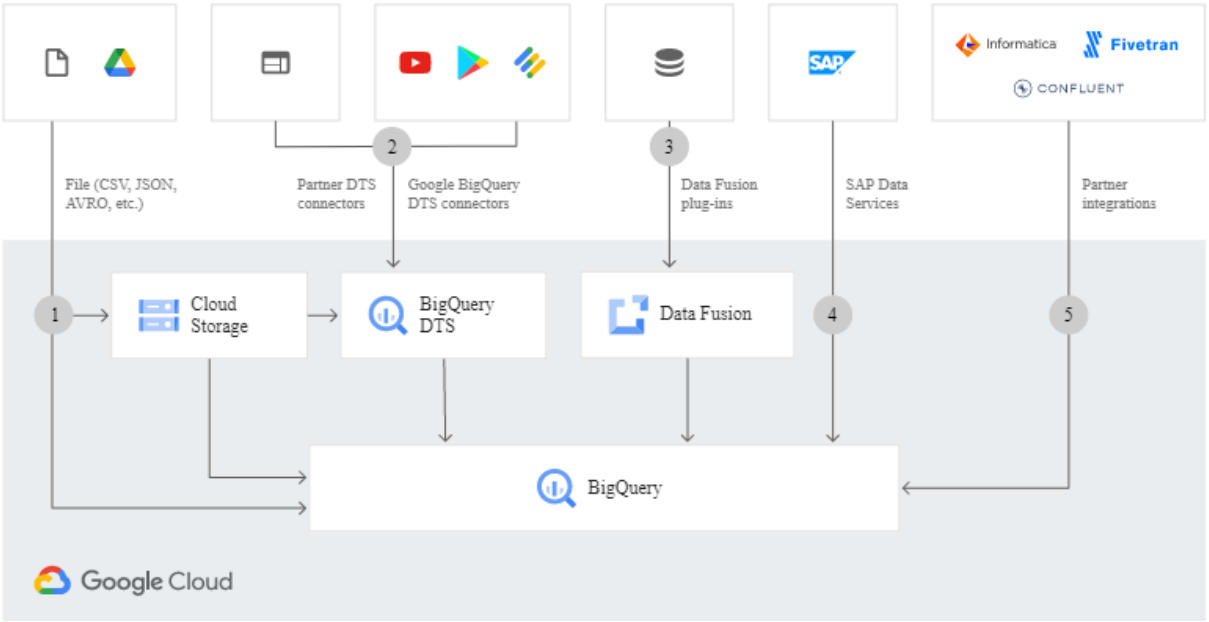
You can use BigQuery jobs to automate and schedule tasks and operations.

With BigQueryML, you can train models with a rich set of algorithms with data already stored in the Cloud. You may perform feature engineering and hyperparameter tuning and export a BigQuery ML model to a Docker image as required.

All other services are useful in ML pipelines, but they aren‘t that easy and ready to use.

Vertex AI is a new API that combines AutoML and AI Platform. You can use both AutoML training and custom training in the same environment.

It obviously has a rich set of features for managing ML pipelines.



For any further detail:

https://cloud.google.com/bigquery-ml/docs/export-model-tutorial

https://cloud.google.com/bigquery/docs/jobs-overview

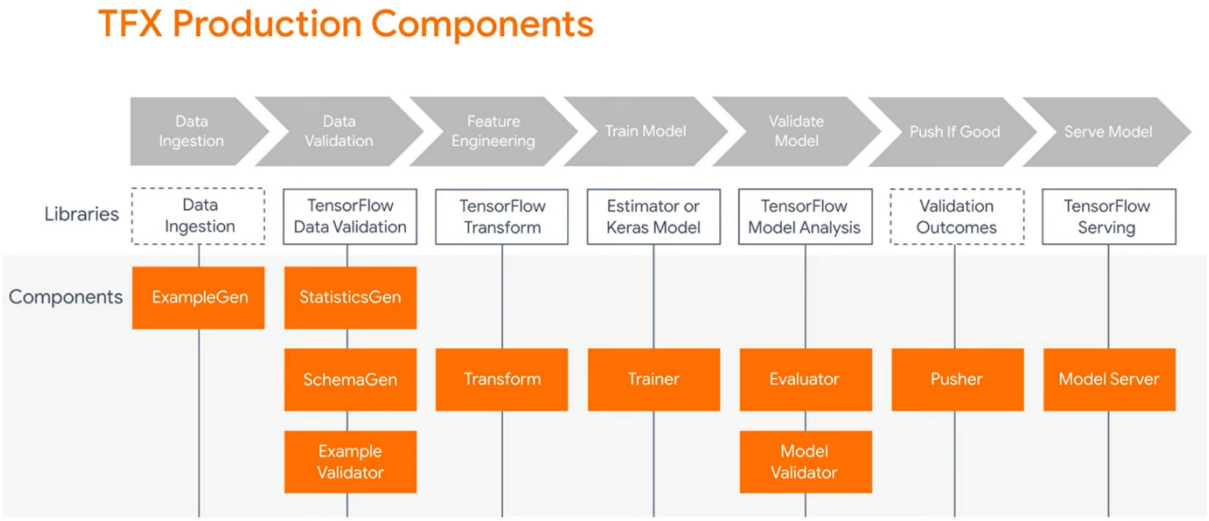
https://cloud.google.com/bigquery/

Question #19 – What is TFX?

Answer #19 - TFX is a platform that allows you to create scalable production ML pipelines for TensorFlow projects, therefore Kubeflow.

It, therefore, allows you to manage the entire life cycle seamlessly from modeling, training and validation, up to production start-up and management of the inference service.

AI Platform manages TFX, under AI Platform and pipelines:



For any further detail:

https://cloud.google.com/ai-platform/pipelines/docs

https://developers.google.com/machine-learning/crash-course/production-ml-systems

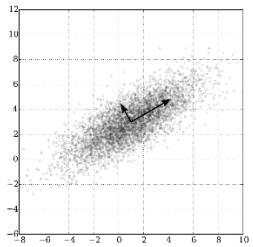
https://www.tensorflow.org/tfx/guide

https://www.youtube.com/watch?v=Mxk4qmO_1B4

Question #20 – What is Principal Component Analysis?

Answer #20 - Principal component analysis is a technique to reduce the number of features by creating new variables obtained from linear combinations or mixes of the original variables, which can then replace them but retain most of the information useful for the model. In addition, the new features are all independent of each other.

The new variables are called principal components.



For any further detail:

https://developers.google.com/machine-learning/crash-course/embeddings/categorical-input-data

https://builtin.com/data-science/step-step-explanation-principal-component-analysis

https://en.wikipedia.org/wiki/Principal_component_analysis

Question #21 – Can we do automatic deployment and serving in BigQuery ML?

Answer #21 – No. In BigQuery ML we can perform below tasks:

A. Exploratory data analysis

B. Feature selection

C. Model building

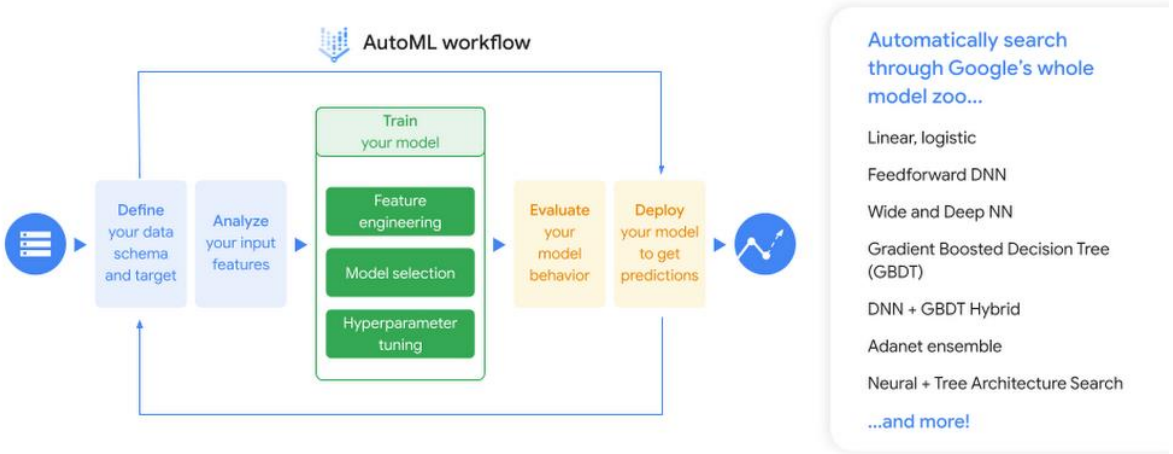
D. Training

E. Hyperparameter tuning

BigQuery is perfect for Analytics. So, exploratory data analysis and feature selection are simple and very easy to perform with the power of SQL and the ability to query petabytes of data.

BigQuery ML offers all other features except automatic deployment and serving.

BigQuery ML can simply export a model (packaged in a container image) to Cloud Storage.



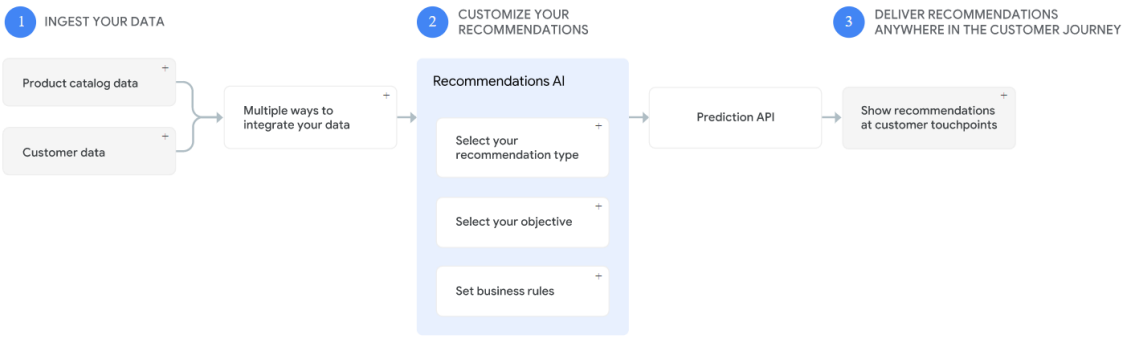
For any further detail:

https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-export-model

https://cloud.google.com/blog/products/data-analytics/automl-tables-now-generally-available-bigquery-ml

Question #22 – What is Recommendations AI?

Answer #22 - Recommendations AI is a ready-to-use service for all the requirements shown in the question. You don't need to create models, tune, train, all that is done by the service with your data. Also, the delivery is automatically done, with high-quality recommendations via web, mobile, email. So, it can be used directly on websites during user sessions.



For any further detail:

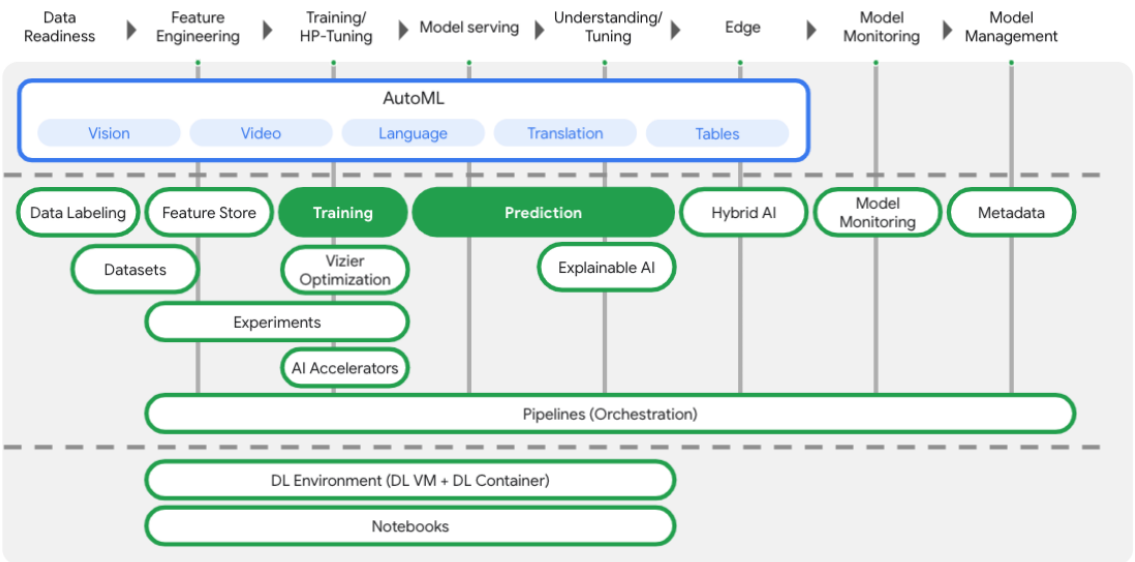
https://cloud.google.com/retail/recommendations-ai/docs/create-models

https://cloud.google.com/recommendations

Question #23 – How to do custom tooling and pipelines in Vertex AI?

Answer #23 - Vertex AI combines AutoML, custom models and ML pipeline management through to production. Vertex AI integrates many GCP ML services, especially AutoML and AI Platform, and includes many different tools to help you in every step of the ML workflow.

So, Vertex AI offers two strategies for model training: AutoML and Personalized training.



Machine learning operations (MLOps) is the practice of using DevOps for machine learning (ML).

DevOps strategies automate the release of code changes and control of systems, resulting in greater security and less time to get systems up and running.

For any further detail:

https://cloud.google.com/vertex-ai/docs

https://cloud.google.com/vertex-ai/docs/pipelines/introduction

https://codelabs.developers.google.com/codelabs/vertex-ai-custom-models#1

Question #24 – When you train your DNN with Tensorflow, your input data does not fit into RAM memory. What can you do in the simplest way?

Answer #24 - Use tf.data.Dataset. The tf.data.Dataset allows you to manage a set of complex elements made up of several inner components.

It is designed to create efficient input pipelines and to iterate over the data for their processing.

These iterations happen in streaming. So, they work even if the input matrix is very large and doesn't fit in memory.

For any further detail:

https://www.tensorflow.org/api_docs/python/tf/data/Dataset

https://www.kaggle.com/jalammar/intro-to-data-input-pipelines-with-tf-data

Question #25 – What is AutoML tables?

Answer #25 - AutoML Tables can select the best model for your needs without having to experiment.

The architectures currently used (they are added at the same time) are:

Linear

Feedforward deep neural network

Gradient Boosted Decision Tree

AdaNet

Ensembles of various model architectures

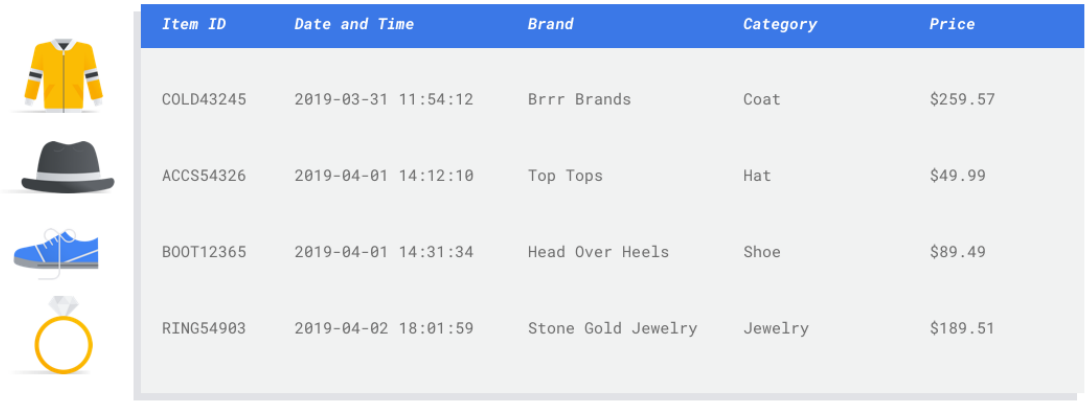
In addition, AutoML Tables automatically performs feature engineering tasks, too, such as:

- Normalization
- Encoding and embeddings for categorical features.

- Timestamp columns management (important in our case)

So, it has special features for time columns: for example, it can correctly split the input data into training, validation and testing.

Vertex AI is a new API that combines AutoML and AI Platform. You can use both AutoML training and custom training in the same environment.



Item ID	Date and Time	Brand	Category	Price
COLD43245	2019-03-31 11:54:12	Brrr Brands	Coat	$259.57
ACCS54326	2019-04-01 14:12:10	Top Tops	Hat	$49.99
BOOT12365	2019-04-01 14:31:34	Head Over Heels	Shoe	$89.49
RING54903	2019-04-02 18:01:59	Stone Gold Jewelry	Jewelry	$189.51

For any further detail:

https://cloud.google.com/automl-tables/docs/features

https://cloud.google.com/vertex-ai/docs/pipelines/introduction

https://cloud.google.com/automl-tables/docs/beginners-guide

Question #26 – What is Generative Adversarial Network (GAN)?

Answer #26 – GAN is a special class of machine learning frameworks used for the automatic generation of facial images. GAN can create new characters from the provided images.

It is also used with photographs and can generate new photos that look authentic.

For any further detail:

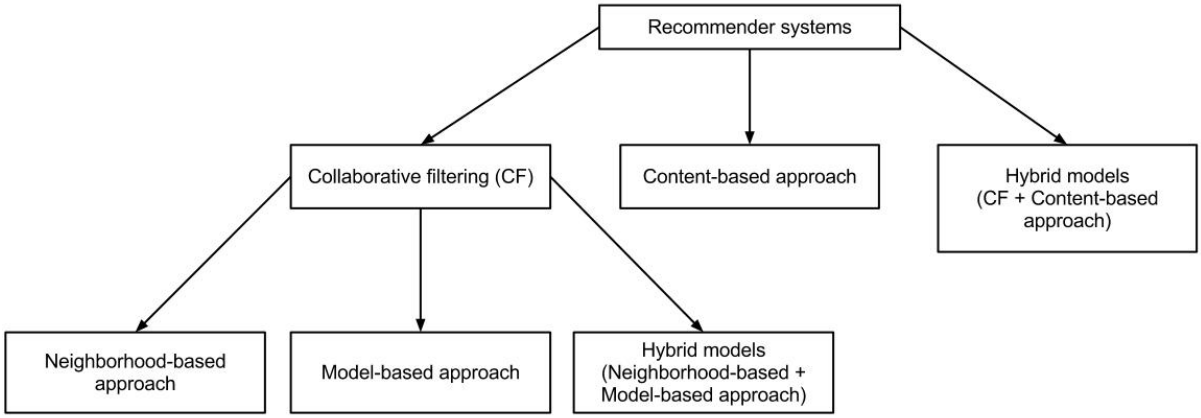
https://en.wikipedia.org/wiki/Generative_adversarial_network

https://developer.nvidia.com/blog/photo-editing-generative-adversarial-networks-2/

Question #27 – What is collaborative filtering?

Answer #27 – Collaborative filtering works on the idea that a user may like the same things of the people with similar profiles and preferences.

So, exploiting the choices of other users, the recommendation system makes a guess and can advise people on things not yet been rated by them

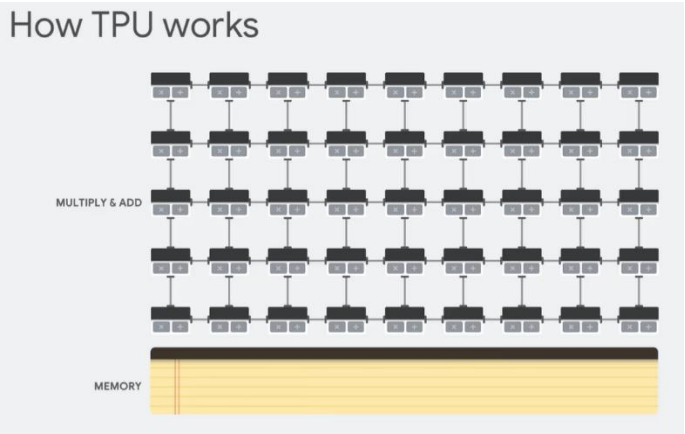


Question #28 – How to make best use of TPU?

Answer #28 – You may use preemptible Cloud TPU (70% cheaper) for your fault-tolerant machine learning workloads.

You may use TPUs in the AI Platform because TensorFlow APIs and custom templates can allow the managed environment to use TPUs and GPUs using scale tiers.

You may optimize your workload using the Profiler with TensorBoard. TensorBoard is a visual tool for ML experimentation for Tensorflow.



For any further detail:

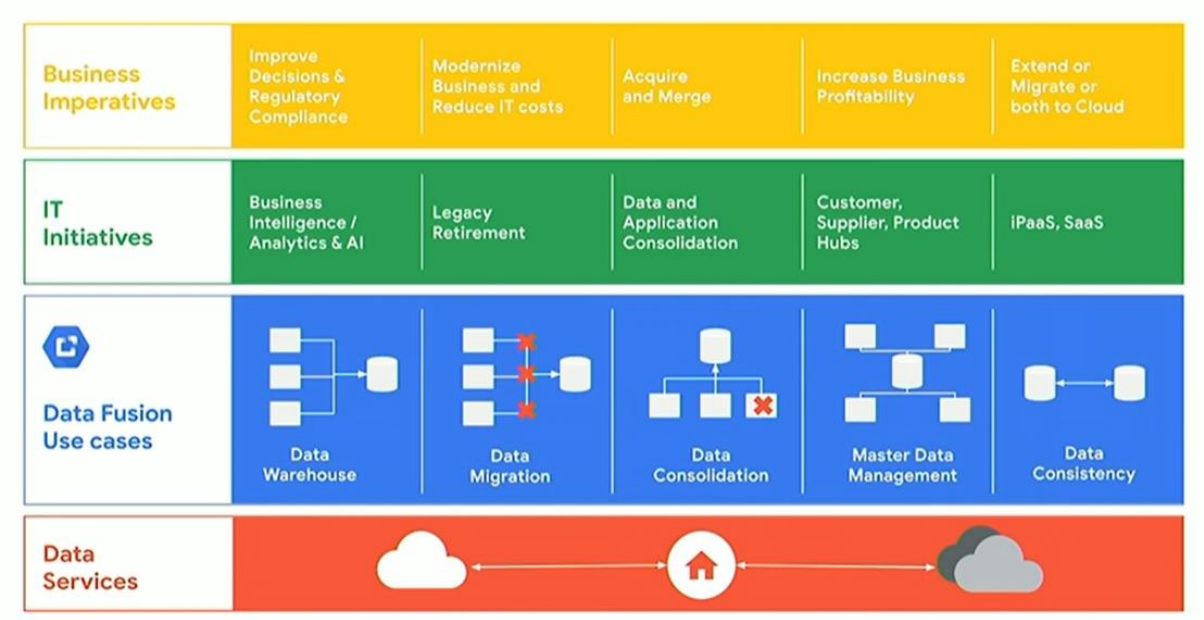
https://storage.googleapis.com/nexttpu/index.html

https://cloud.google.com/ai-platform/training/docs/using-tpus

Question #29 – What is Cloud Data Fusion?

Answer #29 – Cloud Data Fusion is a managed service for quickly building data pipelines and ETL processes. It is based on the open-source CDAP project and therefore is portable to any environment.

It has a visual interface that allows you to create codeless data pipelines as required.



For any further detail:

https://cloud.google.com/data-fusion

https://www.youtube.com/watch?v=kehG0CJw2wo

Question #30 – What are the common encoding techniques you have?

Answer #30 – As below:

- Ordinal Encoding: Ordinal encoding simply creates a correspondence between each unique category with an integer.
- One-Hot Encoding: One-hot encoding creates a sparse matrix with values (0 and 1, see the picture) that indicate the presence (or absence) of each possible value.

Color
Red
Red
Yellow
Green
Yellow

→

Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1
0	1	0

- Embeddings: Embeddings are often used with texts and in Natural Language Processing (NLP) and address the problem of complex categories linked together.

- Feature Crosses: Feature crosses creates a new feature created by joining or multiplying multiple variables to add further predictive capabilities, such as transforming the geographic location of properties into a region of interest.

For any further detail:

https://developers.google.com/machine-learning/crash-course/embeddings/categorical-input-data

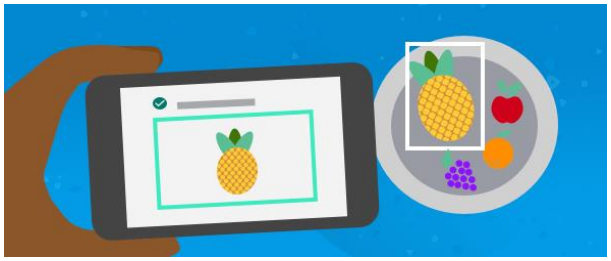
https://developers.google.com/machine-learning/crash-course/feature-crosses/crossing-one-hot-vectors

https://www.kaggle.com/alexisbcook/categorical-variables

Question #31 – When to use AutoML vision?

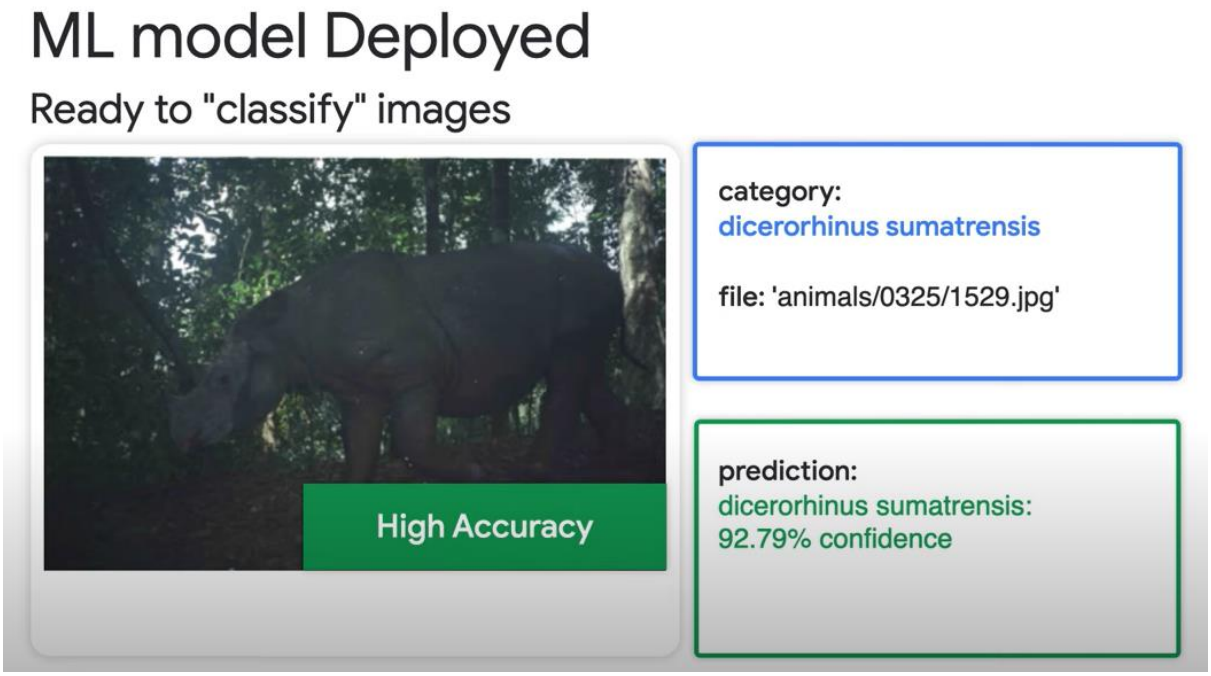
Answer #31 – **Vision AI** uses pre-trained models trained by Google. This is powerful, but sometimes not enough.

AutoML Vision Edge is for local devices.



Video AI manages videos, not pictures. It can extract metadata from any streaming video, get insights in a far shorter time, and let trigger events.

But **AutoML Vision** lets you train models to classify your images with your own characteristics and labels. So, you can tailor your work as you want.



For any further detail:

https://cloud.google.com/vision/automl/docs/edge-quickstart

https://cloud.google.com/vision/automl/docs/beginners-guide

https://cloud.google.com/natural-language/

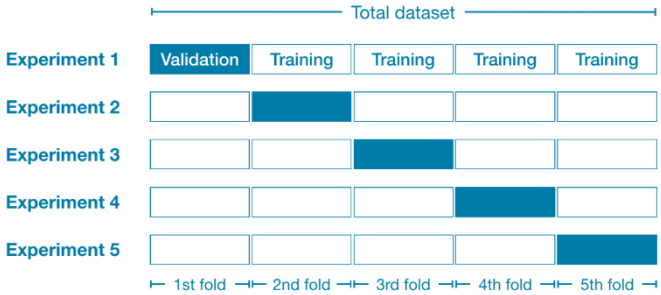
https://cloud.google.com/automl

https://www.youtube.com/watch?v=hUzODH3uGg0

Question #32 – What is Cross-Validation?

Answer #32 – Cross-validation involves running our modeling process on various subsets of data, called “folds”.

Obviously, this creates a computational load. Therefore, it can be prohibitive in very large datasets, but it is great when you have small datasets.



For any further detail:

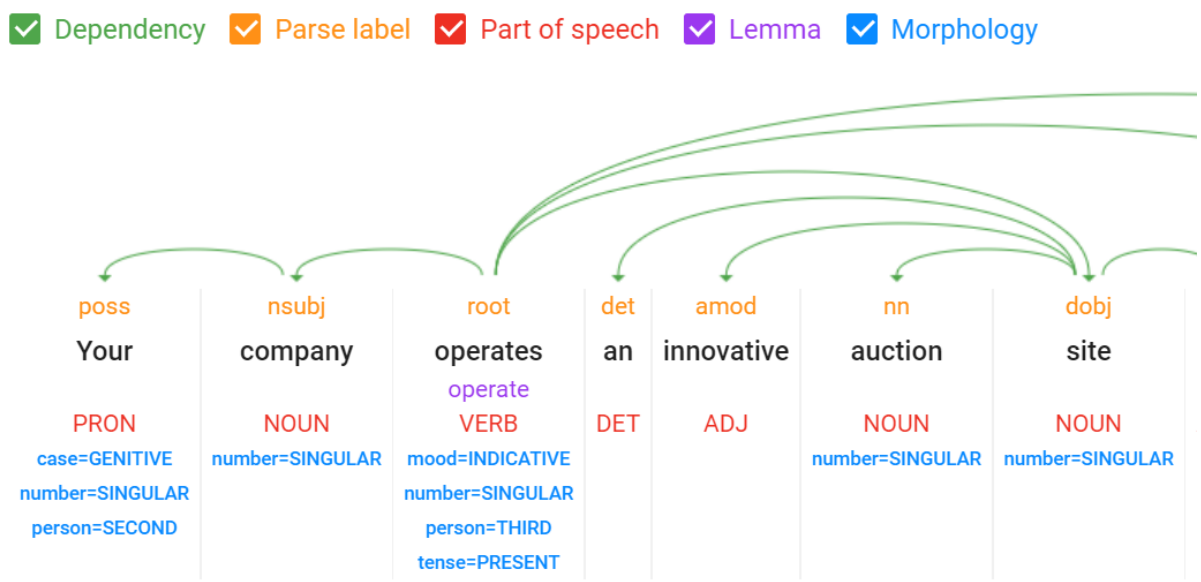
https://developers.google.com/machine-learning/glossary?hl=en#cross-validation

https://www.kaggle.com/alexisbcook/cross-validation

Question #33 – What are the NLP related products/language GCP offers?

Answer #33 –

- **Speech to Text** can convert voice to written text.
- **Cloud Natural Language API** can understand text meanings such as syntax, feelings, content, entities and can create classifications.
- **Text to Speech** can convert written text to voice.
- **Speech Synthesis Markup Language (SSML)** is not a service but a language used in Text-to-Speech requests. It gives you the ability to indicate how you want to format the audio, pauses, how to read acronyms, dates, times, abbreviations and so on. Really, it is useful for getting closer to human dialogue.



For any further detail:

https://cloud.google.com/speech-to-text

https://cloud.google.com/text-to-speech/docs/basics

https://cloud.google.com/text-to-speech/docs/ssml

https://cloud.google.com/natural-language/

Question #34 – Your model has many independent variables. You discovered that you could not achieve good results because many variables are correlated. Which techniques or algorithms did he advise to use?

Answer #34 – If you have many independent variables, some of which are correlated with each other. You have multicollinearity; therefore, you cannot use classical linear regression.

Partial Least Squares and Principal components create new variables that are uncorrelated.

Partial Least Squares method uses projected new variables using functions.

The main **PCA** components reduce the variables while maintaining their variance. Hence, the amount of variability contained in the original characteristics.

Multivariate regression finds out ways to explain how different elements in variables react together to changes.

For any further detail:

https://towardsdatascience.com/partial-least-squares-f4e6714452a

https://en.wikipedia.org/wiki/Partial_least_squares_regression

https://towardsdatascience.com/maximum-likelihood-estimation-984af2dcfcac

https://en.wikipedia.org/wiki/Partial_least_squares_regression

https://www.mygreatlearning.com/blog/introduction-to-multivariate-regression/

https://colab.research.google.com/github/kaustubholpadkar/Predicting-House-Price-using-Multivariate-Linear-Regression/blob/master/Multivariate_Linear_Regression_Python.ipynb

https://en.wikipedia.org/wiki/Polynomial_regression

Question #35 – You are using an AI Platform, and you are working with a series of demanding training jobs. So, you want to use TPUs instead of CPUs. You are not using Docker images or custom containers. What is the simplest configuration to indicate if you do not have particular needs to customize in the YAML configuration file?

Answer #35 – AI Platform lets you perform distributed training and serving with accelerators (TPUs and GPUs).

You usually must specify the number and types of machines you need for master and worker VMs. But you can also use scale tiers that are predefined cluster specifications.

In our case,

scale-tier=BASIC_TPU

covers all the given requirements.

For any further detail:

https://cloud.google.com/ai-platform/training/docs/machine-types#scale_tiers

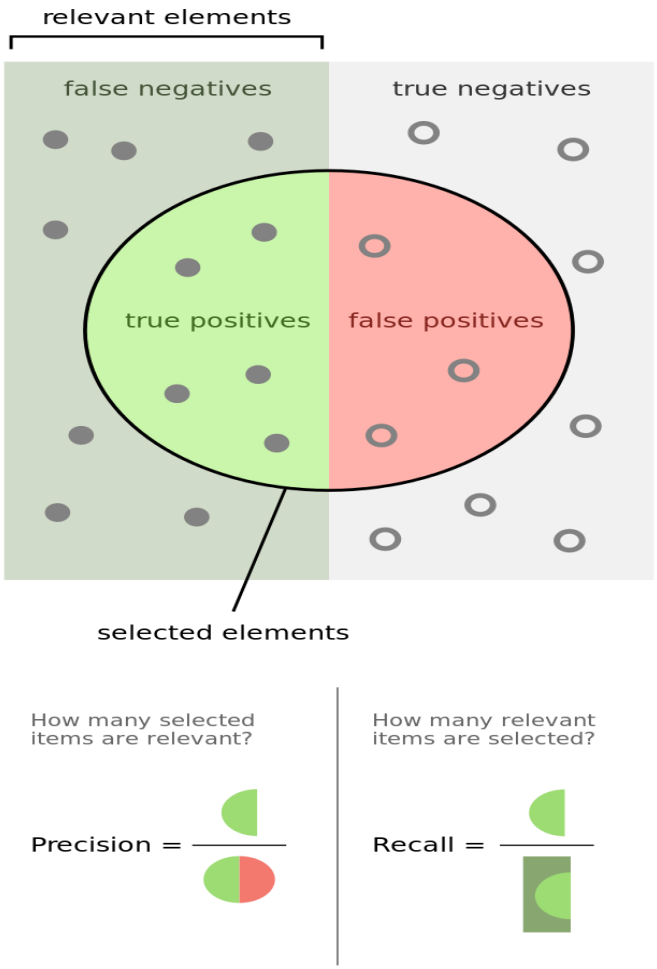
https://cloud.google.com/ai-platform/training/docs

https://cloud.google.com/ai-platform/training/docs/using-tpus#configuring_a_custom_tpu_machine

https://cloud.google.com/tpu/docs/tpus

Question #36 – What is Recall?

Answer #36 – Recall indicates how true positives were recalled (found).



For any further detail:

https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall

https://en.wikipedia.org/wiki/F-score

https://en.wikipedia.org/wiki/Precision_and_recall#/media/File:Precisionrecall.svg

Question #37 – What are the embedding techniques you have in NLP?

Answer #37 –

A **Count Vector** gives a matrix with the count of every single word in every example. 0 if no occurrence. It is okay for small vocabularies.

TF-IDF vectorization counts words in the entire experiment, not a single example or sentence.

Co-Occurrence Matrix puts together words that occur together. So, it is more useful for text understanding.

For any further detail:

https://developers.google.com/machine-learning/crash-course/embeddings/categorical-input-data

https://developers.google.com/machine-learning/crash-course/feature-crosses/crossing-one-hot-vectors

https://www.analyticsvidhya.com/blog/2017/06/word-embeddings-count-word2veec/

https://towardsdatascience.com/5-things-you-should-know-about-covariance-26b12a0516f1

Question #38 – What is CoVariance Matrix?

Answer #38 – CoVariance Matrices are square matrices with the covariance between each pair of elements. It measures how much the change of one with respect to another is related.

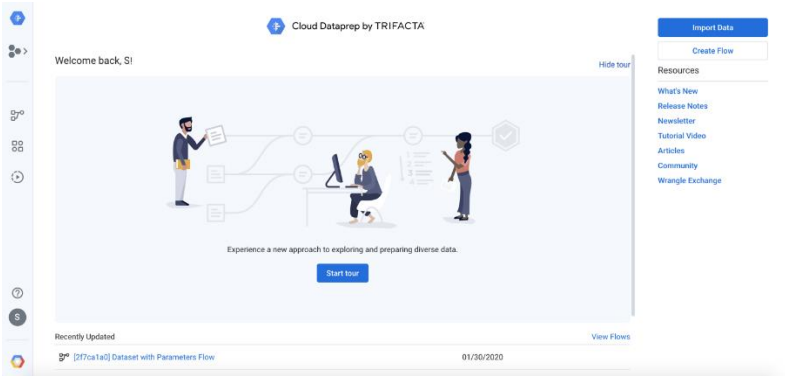
$$\begin{array}{c} \\ x \\ y \end{array} \begin{array}{c} \begin{array}{cc} x & y \end{array} \\ \begin{bmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix} \end{array}$$

For any further detail:

https://www.wikiwand.com/en/Covariance_matrix

Question #39 – What is Dataprep?

Answer #39 – Dataprep is a serverless service that lets you examine clean and correct structured and unstructured data. So, it is fully compliant with our requirements.



For any further detail:

https://cloud.google.com/dataprep

https://docs.trifacta.com/display/dp/

https://developers.google.com/machine-learning/crash-course/representation/cleaning-data

Question #40 – When you use tf.data.Iterator?

Answer #40 – tf.data.Iterator is used for enumerating elements in a Dataset, using Tensorflow API, so it is not suitable for accessing tabular data.

Question #41 – What is Anscombe Quartet problem?

Answer #41 – Anscombe Quartet problem says:

- Not linear relation between independent and dependent variables
- Outliers that change the result

As you may see in the following picture, you may have data without a linear relationship between X and Y that gives you good statistics.