**Question** **1.** **Cloud** **Computing** **(20** **points)**

a. [20 points] A company is trying to decide if it should use Cloud Computing for all of its computing service needs rather than using its own machines in a data centre. As background, this company runs its own web applications, databases and carries out data analysis, machine learning and other ad hoc computing functions.

You have been asked to provide a brief summary of the advantages and disadvantages of using Cloud Computing versus owning and running its own infrastructure and services. These should be stated from technical, economic, security and efficiency perspectives.

*The advantages to using cloud computing is that:*

* *The company will not need to buy or maintain infrastructure that is necessary to manage or maintain web servers and databases. By relying on cloud computing, they will only need to pay for services when being used (or by some subscription periodically). The responsibility to managing and paying for resources or servers will not have to fall on them.*
* *It provides security for resources stored by using various advanced features, such as Key Management Systems, cryptography, and Identify Access Management systems. There are various security measures provided that can help ensure that any private or confidential resources are not compromised by any unauthorized access.*
* *It allows easy access and availability for users regardless of where they are. As long as users are able to connect to these cloud services, they can easily access their resources.*

*The disadvantages to using cloud computing is that:*

* *By relying on cloud services, it is likely that the company could also be using the same services as their competitors. By relying on the same strategy, they may not be able to gain any competitive advantages over them if they know what they are provided.*
* *The services provided by cloud computing may not be as advanced or flexible as what the company is currently managing. If the company chose to move to cloud computing, and migrate their resources to the cloud, there is a chance that they may lose some of their features or attributes (for example, their machine learning functions may not be able to correctly output some predictions due to dependencies) that are not supported by these cloud services.*
* *A strong internet connection (or any connection at all) is required to access these cloud services (compared to company-owned infrastructure, which does not require users to connect to the Internet to connect to it). If the company were located at some place (for example, overseas) with poor or no connection, it would be very unlikely for them to access any of their resources on the cloud.*

**Question** **2.** **Virtual** **Machines** **and** **Containers** **(20** **points)**

a. [10 points] Describe what virtualisation is and describe the characteristic attributes of the different types of virtualisation (Language, Operating System and Hardware).

*Virtualization is a process in which users create a self-contained virtual environment that acts like a computer. This helps simulate software-based resources of what a computer would actually have. This can also help partition resources of physical hardware and servers into multiple virtual machines to share the workload and processes.*

*Language level virtualisation: the VM is created based on some language, that helps us to run some applications based on the language.*

*OS level virtualisation: is a type of virtualization technology which works on the OS layer. For example, if we have the host operating system in a physical computer but want to install another operating system as a virtual machine. And we will have separate applications under different operating systems.*

*Hardware level virtualisation: help to choose what is the resource for the VM (i.e CP, Memory, Network Interfaces or Disks)*

b. [10 points] Describe what containers are with reference to Docker and discuss their similarities and differences from operating system virtualisation perspective as provided by VirtualBox.

**Question** **3.** **Storage** **(10** **points)**

a. [5 points] Describe what S3 is and discuss its similarities and differences with the data store DynamoDB.

*Simple Storage Service (S3) is an AWS cloud service used for object storage. By using S3, buckets are provided to users to store file objects as data and metadata. Users would also be able to determine the permission controls and authentication do determine which users are allowed or not allowed to enact specific actions with the resources in that bucket.*

*Similarities between S3 and DynamoDB is that they are both storage-based services, and have similar features such as highly scalability and availability. The differences between them is that S3 is more useful for storing media files such as images, videos, sound files etc., while DynamoDB focuses on storing document-based files. Another difference is that S3 allows versioning (meaning different versions of the bucket is created every time a new file object is uploaded to it), while DynamoDB does not support it.*

b. [5 points] You are asked to store data about music albums in a DynamoDB table. For each album, you need to record the title of the album and the artist name. Describe the commands you would use to create a table to store such information and write an entry to that table in DynamoDB.

*Create a DynamoDB table to store details of the various albums (as well as define the attributes of the table’s entries). Note that the key-schema determines which attributes of the entry is the partitions key (HASH) and the sort key (RANGE):*

*aws dynamodb create-table –table-name Albums –atribute-definitions AttributeName=Artist,AttributeType=S AttributeName=Title,AttributeType=S –key-schema AttributeName=Artist,KeyType=HASH AttributeName=Title,KeyType=RANGE*

*To write an entry, the following command can be used (assuming the details of the album is already known):*

*aws dynamodb put-item –table-name Albums -item {“Artist”: <Artist’s Name>, “Title”: <Title of the album>}*

**Question** **4.** **Identity** **and** **Access** **Management** **(20** **points)**

a. [20 points] An organisation has 5 departments and has separated out each of the IAM users into separate groups using paths following the pattern companybucket/department1/\*, companybucket/department2/\*, companybucket /department3/\* etc.

Their IAM account names follow the patternuser@department1.company.com, user@department2.company.cometc.

You are tasked with securing a bucket that contains a folder for each of 5 departments in an organisation. Only people within a department can write to their own folder. Everyone can read from all folders.

Discuss the principles that you would use to create a policy that would achieve this objective. Write the policy as a JSON file that you would use.

Note: you can have individual statements for each department.

*The principle that this policy would focus on would be AWS Resources. This policy is used to control access and permissions on who is allowed to view or edit resources within the bucket that the policy is applied to. In this case, the policy should deny edit permissions to users that do not satisfy its conditions (that is, their account name does not have the name of the policy’s department).*

*Assume that the only objective that this policy has is to prevent users that are outside of the specified department to not be able to write any new objects to that folder within the bucket according to the question. Editing and deleting file objects will not be considered. Assume that when the question says everyone can read from all folders, this only refers to anyone in any department within the company (only they are allowed to view the resources in the bucket).*

*The following policies in JSON format show an example of what the policies for all departments could be displayed as:*

*{*

*“Version”: “2022-10-26”,*

*“Statement”: [{*

*“Sid”: “AllowDepartment1GroupPutObject”,*

*“Effect”: “Allow”,*

*“Action”: “s3:PutObject”,*

*“Resource”: “companybucket/department1/\*”,*

*“Condition”: [*

*“StringLike”: “\*@department1.company.com”,*

*]*

*},*

*{*

*“Sid”: “AllowDepartment2GroupPutObject”,*

*“Effect”: “Allow”,*

*“Action”: “s3:PutObject”,*

*“Resource”: “companybucket/department2/\*”,*

*“Condition”: [*

*“StringLike”: “\*@department2.company.com”,*

*]*

*},*

*{*

*“Sid”: “AllowDepartment3GroupPutObject”,*

*“Effect”: “Allow”,*

*“Action”: “s3:PutObject”,*

*“Resource”: “companybucket/department3/\*”,*

*“Condition”: [*

*“StringLike”: “\*@department3.company.com”,*

*]*

*},*

*{*

*“Sid”: “AllowDepartment4GroupPutObject”,*

*“Effect”: “Allow”,*

*“Action”: “s3:PutObject”,*

*“Resource”: “companybucket/department4/\*”,*

*“Condition”: [*

*“StringLike”: “\*@department4.company.com”,*

*]*

*},*

*{*

*“Sid”: “AllowDepartment5GroupPutObject”,*

*“Effect”: “Allow”,*

*“Action”: “s3:PutObject”,*

*“Resource”: “companybucket/department5/\*”,*

*“Condition”: [*

*“StringLike”: “\*@department5.company.com”,*

*]*

*},*

*{*

*“Sid”: “DenyUsersOutsideCompany”,*

*“Effect”: “Deny”,*

*“Action”: “s3:\*”,*

*“Resource”: “companybucket/\*”,*

*“Condition” : [*

*“StringNotLike”: “\*.company.com”,*

*]*

*}]*

*}*

**Question** **5.** **Networking,** **Network** **Address** **Translation** **(10** **points)**

a. [10 points] Discuss the reasons why you would use Application Load Balancing and how this would be set up to load balance a Python Django application. Specifically, describe the configuration of the Listener and Target Group running the Python Django application.

**Question** **6.** **DevOps** **(10** **points)**

a. [10 points] You have been asked to set up the ability to automatically build and test code updated in a repository. Discuss which AWS service you would use to do this, and the steps involved in getting it to work. What information would be needed in the configuration file for this service?

Chef.

**Question** **7.** **Machine** **Learning** **(10** **points)**

a. [7 points] Describe the 3 different types of Machine Learning Models that AWS Machine Learning allows you to use and describe examples of the types of questions you could answer with each one.

*The three types of Machine Learning Models are as follows:*

* 1. *Binary Classification - classifiers an input and predicts if it is a category or another contradicting category*
     1. *Will it be raining or not tomorrow?*
     2. *Is this person in the picture a human or not?*
     3. *According to their resume, did this person work in retail or not?*
  2. *Multiclass Classification - classifies an input and predicts what category it belongs to from a groups of various categories*
     1. *What will the weather be tomorrow? Raining, sunny, or cloudy?*
     2. *What kind of animal is in this picture? A dog, cat or bird?*
     3. *According to their resume, what kind of job did this person work as? Retail, manufacturing, or sales?*
  3. *Regression - predict a value by relying on coefficients that can be used to determine the correlation between the features of the model and the target response (output to be returned)*
     1. *What will the temperature be tomorrow?*
     2. *How tall is the person in the picture?*
     3. *What will the price of the house be in 2 years time?*

b. [3 points] A dog recognition program recognises 10 dogs in a picture of 14 dogs and some cats. Of the 10 dogs, 7 are true positives and 3 are false positives.

i. How many actual dogs did the recognition program recognise?

*True positives = 7*

ii. What is the precision of the program?

*Precision = True positives / (True positives + False positives)*

*= 7 / (7 + 3)*

*= 0.7*

iii. What is the recall of the program?

*Recall = True positives / True number*

*= 7 / 14*

*= 0.5*

**Question** **1.** **Cloud** **Computing** **(20** **points)**

a. [20 points] Describe the concepts of *vertical* and *horizontal* scale. Describe 2 different ways in which you could scale a web application horizontally. Describe a potential architecture to scale a database to handle the scaling out of web servers.

**Question** **2.** **Virtual** **Machines** **and** **Containers** **(20** **points)**

a. [10 points] Describe what *virtualisation* is. Describe the characteristic attributes of the different types of virtualisation (Language, Operating System and Hardware).

b. [10 points] Describe what *containers* are with reference to Docker. Discuss their similarities and differences from an operating system virtualisation perspective as provided by VirtualBox or VMware.

**Question** **3.** **Storage** **(10** **points)**

a. [ 10 points] Describe what *S3* is and describe its “eventual consistency” mechanism. What are the potential considerations if you are writing a multi-user application that uses S3? What other technology could you use if you want to avoid the problems of eventual consistency?

**Question** **4.** **Identity** **and** **Access** **Management** **(20** **points)**

a. [20 points] A medium sized company has users that belong in different departments and perform different functions. The company has implemented a policy of document access that is specific to the job that a person does and their level in the organisation.

Describe how you could use *AWS* *IAM* to provide authorization and authentication in this organisation to access to documents, as well as access and perform actions using applications.

**Question** **5.** **Networking,** **Network** **Address** **Translation** **(20** **points)**

a. [ 10 points] When an EC2 instance is created in AWS, it is assigned to a region and a Virtual Private Cloud (VPC).

Describe how network addresses are allocated to a VPC and sub-networked when an EC2 instance is created. How is the EC2 connected to other machines and to the Internet?

b. [10 Points] Describe 2 ways in which security is implemented in AWS networks at the network level. What are the similarities and differences between these 2 security implementations?

**Question** **6.** **DevOps** **(10** **points)**

a. [10 points] What is *DevOps*? Describe how you would implement the automation of creation of machines, configuration of software and deployment of application programs using AWS.

**1.** **(20** **points)**

The creative designers of a marketing and video editing company in Melbourne use a commercial web-based application for converting videos to different formats. In addition, one of the main advantages of this software is the ability to publish those videos in popular social media networks such as YouTube, Instagram, Twitter, etc. This allows the user to convert and edit a video and then publish the same video at the same time to all different social media that the user selects in the software. This is particularly useful for marketing campaigns were usually a video is published in different social media networks. In the past, before using that software they used to spend around 5-10 hours to convert a single video to different formats, resolutions, etc. and upload them to all different social media. Now using this software this process is reduced to only 45mins to 1 hour. Since they started using this software, they have to pay a monthly subscription fee of 25.000 AUD per month or around 300.000 AUD per year regardless of the number of videos they process per year. Despite all the benefits from this software, the CTO of the company believes they should pay a software developer to replicate this web-based software instead of keep paying those high fees. While he knows building software is expensive, he also believes there must be a way of designing this new software to reduce the cost and make it profitable in the long term. You as a Software Developer and Cloud Computing expert are hired for this task.

1.1 **(10** **points)** You initially think on using EC2 instances for processing. Briefly explain what other AWS technologies you would use to replicate the web-based software functionalities and what architecture you would use to satisfy the needs of the client using such technologies.

S3?

1.2 **(10** **points)** The company currently creates on average 50 videos per month and it’s not likely to increase the demand of clients for the next couple of years. Take also into account that there are less than 20 people that will be using the system. Discuss how you could change the previous solution using now a full serverless architecture and what benefits and disadvantages would have this new architecture over the previous architecture using EC2 instances.

Lambda? Microservice Architecture?

*By using serverless architecture, the company will not need to maintain or manage any software infrastructure or servers when they need to upload or process the videos. Regarding the size of the company and the fact that they currently create an average of 50 videos per month, the workload that is required to create, refine, and upload a large number of videos can be too much if this was all done manually. It is also likely that the demand for more videos would be introduced after the next couple of years. The company could rely on DevOps (which can involve serverless computing) to allow them to automate the process of creating these videos when necessary. With serverless computing, they would be able to pass the job from creating videos to uploading to some repository or social media through some service or tool that can do all of that without any manual effort.*

*An advantage to serverless computing is that with EC2, the company would need to create and maintain multiple instances or virtual machines to run code that handles the processes. Serverless computing will only need to run code when necessary; it is event-driven. Because of that, it can be seen as a pay-on-demand service, where you only need to pay as the automation process runs, instead of paying to continuously manage and store instances to run even when inactive.*

*A disadvantage to serverless computing falls on the fact that this does not rely on a server, meaning there could be some potential dangers to security and privacy. EC2 instances and virtual machines run on servers that could require authorization before users are able to access them. Serverless computing does not provide any authorization checks; the code is simply run when called automatically…*

**2.** **(20** **points)**

A professor in Computer Science at UWA has data from all labs, mid-term exam and final exam marks for the last two years of every student in csv format. He first noticed that in previous years those students that did very well in the labs and mid-term exam for a particular unit got at least 70% marks in the final exam. This year, the professor only has access to the labs and mid-term exam marks in csv format. However, he thinks he can predict the outcome of the final exam for students before marking the final exam.

2.1 **(10** **points)** How would you use AWS technologies to prove (or disprove) his theory and help the professor predicting the marks for the current students this year?

*The professor could use AI-based technologies provided on AWS to create a model to predict the final exam marks of students based on the marks they got for the other assessments. An example of this would be using SageMaker, a service that helps create an environment that would be capable of carrying out machine learning tasks (it can create and execute programs on Jupyter, which can run these machine learning processes).*

*To make the model accurate with predicting these final marks, the professor will have to first understand what sort of data they will need to provide to achieve that goal. In this case, it would be the students’ marks for their labs and the mid-semester test.*

*When the data has been gathered, the professor needs to split the data up into a training set (to train the machine learning model), and some calibration / validation set (to determine if the model after training has achieved a desirable accuracy of predicting final exam marks). A commonly used ratio would be 70:30 or 80:20 for training and calibration respectfully.*

*The predictions of the machine learning model when being fed the data could depend on what type of model is being used. Let us say some regression model that accepts two continuous numerical values is used to predict the final exam marks. After an iterative process of training, and confirming its accuracy with the testing or validation, the professor can determine whether the model is trustworthy enough to correctly predict the marks of students without significant errors. If so, they could deploy this model into production to be ready for use.*

2.2 **(10** **points)** What approach would you use to assess the data models and help with the data analysis of the results?

*There are some evaluation measures that can be used to assess the data models and analyse their performances. Following from the previous answer of a regression model with multiple variables, perhaps the predictions of the model could be plotted to visualize the performance, such as the RMSE (root mean squared error) to measure the distance between the predicted final marks, to the actual final marks (the ground truths of the data). By recording the error (or residuals) between the two, they can be used to visualize how accurate the regression model is in predicting the final exam marks (checks if the regression line fits into the distribution of the ground truths).*

*If a single model was tested, the professor may compare the results of the performance analysis with some criteria (say the RMSE is less than some threshold to be trustworthy). If it achieves this check, the professor can trust that the model is able to predict the correct value most of the time. If not, perhaps they should re-evaluate the model and consider choosing a different type of machine learning model to get better predictions. If there were various models tested for predicting, the professor could compare their performance measures and plots together to determine which one performs better or is more reliable in predicting the correct final marks.*

**3.** **(10** **points)**

A software developer wants to create a Drobox-like application for Photographers where they could seamlessly upload pictures from a local machine to the cloud and synchronise the data across multiple devices (Computer, Mobile Phones, etc.) for authenticated users. For every image uploaded by the users, she wants to save the metadata related to the file. Given the files uploaded by users are very important, she wants to allow restoring such files even if users delete the images on purpose. In addition, the software developer wants to automatically create labels for the uploaded images so they can perform searches of images based on the generated labels.

Describe your approach using AWS technologies to create that application.

**4.** **(20** **points)**

A MedTech company in Perth has created a novel algorithm for detecting heart diseases from **X-rays** ***and*** ***CT*** ***scans*** ***(medical*** ***images)*** using machine learning and computer vision. The company now wants to commercialise it as a Software-as-a-service (SaaS) product. Initially, this product should allow authorised radiologist working at the Royal Perth Hospital to upload such images. Then, once the algorithm can process the image (it could take a couple of minutes) it should retrieve the prediction. The solution should be scalable so it can be used in other hospitals in the future.

Give 10 recommendations to the company on how they could use AWS to create this platform. **Provide** **your** **answer** **as** **a** **list** **of** **bullet** **points.**

**5.** **(20** **points)**

You are a consultant who has been asked to write a report for a rapidly growing pet food company (PFC), based in Perth who would like to upgrade their systems to cope with the increasing global demand for their products. The company has rented space in a data centre for their systems that are a mix of Windows and Linux Servers, networking and firewall equipment and a range of storage devices (SANs, NAS, disk drives in servers). The company uses a range of software that they have purchased over the years that run on these machines. The functionality covered is everything from finance, sales and manufacturing to online product sales through their website. PFC have received venture capital funding and want to expand into new markets, especially Asia and China in particular. To do this, they will need to be able to use modern systems that scale and operate at global scale.

Provide 5 pros and 5 cons of moving infrastructure from being on-premises to cloud based. Consider this in the context of

global expansion and resilience. **Provide** **your** **answer** **as** **a** **list** **of** **bullet** **points.**

**Weekly Questions:**

Week 7

[17] [10 points] Discuss the reasons why you would use Application Load Balancing and how this would be set up to load balance a Python Django application. Specifically, describe the configuration of the Listener and Target Group running the Python Django application.

Load balancing efficiently distributes incoming network traffic across a group of backend server. It helps scaling servers up or down based on the traffic. Reason for implementing a load balancer:

1) Distributes client requests or network load efficiently across multiple servers

2) Ensures high availability and reliability by sending requests only to servers that are online

3) Provides the flexibility to add or subtract servers as demand dictates

4) Improves the performance and availability of applications, websites, databases, and other computing resources.

5) Helps process user requests quickly and accurately.

6) Allows horizontal scale

a. Greater demand, add more instances

b. Less demand, remove instances

7) Allows for update of code on instances without interrupting service

8) Can handle termination of SSL/TLS

a. Less load on instances

b. Simpler code on instances

9) Adds a layer of security

a. Instances are not directly accessible from the Internet

When using an Application Load Balancer, the set up would be the following:

1) Create the Load Balancer

2) Create Target group using the same VPC of the instance. The targets in the target group should be the id of the instance/s created.

3) Create Listener. In the configuration this should include the ARN of the Load Balancer created, protocol should be HTTP and port 80. Also, the actions or rules of the listener should be ‘forward’ so it routes requests to its target group, which will be specified in the configuration by the ARN of the target group.

4) After installing apache in the instances and accessing the the load balancer DNS name should provide the Python Django app.

[18] [10 Points] Describe 2 ways in which security is implemented in AWS networks at the network level. What are the similarities and differences between these 2 security implementations?

1. Security groups: This is basically a firewall for each instance. You define the security groups within the VPC and then assign them to instances. You can have one security group that allows (say) HTTP or HTTPS to the instance; and then apply that to many instances.
2. Network Access Control Lists: You apply these at a subnet level and they are good for large-scale filtering: For example, I want this network to talk to this other network. While you \*can\* use it to filter traffic between individual instances in my judgment you shouldn’t – use security groups for that.

They’re both similar in the way that when used, there’s a security implementation for the instance access. They’re differences are:

| **Security Group** | **Network ACL** |
| --- | --- |
| Operates at the instance level (first layer of defense) | Operates at the subnet level (second layer of defense) |
| Supports allow rules only | Supports allow and deny rules |
| Is stateful: Return traffic is automatically allowed, regardless of any rules | Is stateless: Return traffic must be explicitly allowed |
| All rules are evaluated before deciding to allow traffic | We process rules in order when deciding to allow or deny traffic |
| Specific to the instance it is assigned to | Automatically applies to all traffic in the subnet it is assigned to; therefore applies to all instances in that subnet |

Week 8

[19] [10 points] When an EC2 instance is created in AWS, it is assigned to a region and a Virtual Private Cloud (VPC). Describe how network addresses are allocated to a VPC and sub-networked when an EC2 instance is created. How is the EC2 connected to other machines and to the Internet?

[20] [10 points] What is DevOps and describe how you would implement the automation of creation of machines, configuration of software and deployment of application programs using AWS.

Week 9

[21] [10 points] Describe the 3 different types of Machine Learning Models that AWS Machine Learning allows you to use and describe examples of the types of questions you could answer with each one.

Answer is here: <https://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html>

[22] [10 points] Describe the different ways in which a Microservice Architecture can be implemented on AWS and what the benefits of this approach might be. How could authentication and authorisation be handled in this approach?

<https://aws.amazon.com/microservices/>

<https://www.appdynamics.com/topics/benefits-of-microservices>

With a microservices architecture, an application is built as independent components that run each application process as a service. These services communicate via a well-defined interface using lightweight APIs. Services are built for business capabilities and each service performs a single function. Because they are independently run, each service can be updated, deployed, and scaled to meet demand for specific functions of an application.

Microservices can be hosted on Lambda (Serverless), Containers, or can be other Django/Ruby/NodeJS apps

Benefits:

1) Improved productivity: Breaking an application down into smaller autonomous fragments makes it easier to build and maintain. Each service can be developed, deployed, and managed independently, and can utilize different programming languages, different technology, and different software environments based on the needs of each.

2) Better resiliency: Implementing microservice-based architecture adds ease to the process of identifying and resolving the root cause of performance issues. The improved fault isolation offered by individual modules means larger applications remain unaffected by a single failure.

3) Increased scalability: The fact that each service can be written in a different language or technology allows DevOps teams to choose the most appropriate tech stack for each module without concerns about incompatibility.

Individual services can also be scaled independently, and new components can be added without requiring downtime and redeployment of the entire system.

4) Continuous integration/continuous delivery (CI/CD): Continuous integration and continuous delivery are key concepts of both the DevOps philosophy and the agile approach. Microservices architecture allows cross-functional teams to develop, test, problem-solve, deploy, and update services independently, which leads to faster deployment and troubleshooting turnaround times.

5) Optimize business functionality: When the focus is on a specific service versus the entire application, it’s easier to customize the needs of each component to improve business functionality. Working on individual modules allows teams to focus on business capabilities instead of technologies.

6) Agility: Microservices foster an organization of small, independent teams that take ownership of their services. Teams act within a small and well understood context, and are empowered to work more independently and more quickly. This shortens development cycle times. You benefit significantly from the aggregate throughput of the organization.

7) Easy Deployment: Microservices enable continuous integration and continuous delivery, making it easy to try out new ideas and to roll back if something doesn’t work. The low cost of failure enables experimentation, makes it easier to update code, and accelerates time-to-market for new features.

8) Language agnostic

9) Maintains an interface that encapsulates implementation

10) Can use SQS to allow queues to invoke APIs

For Authorisation need to handle authentication and authorisation for the API:

· Use IAM either directly or through API Gateway Resource Policies

· Use Cognito

· Use API keys

· Use an external service through Lambda

Week 10

[23] [20 points] [a] [15 points] You have been asked to create a project plan for the new machine learning model your company has asked you to build. List the main tasks and sub-tasks you would need to complete to create the model on AWS. [b] [5 points] How would you measure the accuracy of the model you created?

a)

* Identify the business objective that the machine learning model is aiming to address.
* Link the data to AWS and prepare it (clean the data, remove redundant features)
* See whether a robust “off the shelf” machine learning solution exists already. If not…
* Split the data into training and test sets.
* Select an appropriate machine learning model for the task at hand, and specify model parameters
* Train the model
* Evaluate the model by looking at performance metrics.
* Retrain model/generate features as required.
* Finalise model and deploy the model into production.

These steps can be performed in Sagemaker (for custom models) or using services like AWS Rekognition, or AWS Comprehend.

b) Various metrics can be used to measure the accuracy of the model depending on what the model is. For classification models (categorical predictions), one can look at the accuracy, precision, recall, and F1 scores. One can also look at the area under the ROC curve. These metrics show different characteristics of the model, and some may be of more interest depending on the application of the model (e.g. high recall for a cancer diagnosis tool). For regression models (scalar predictions) one can use RMSE (root mean squared error) or other metrics (eg. MAE) which examines the distance from the prediction and ground truth.

It is important to contextualise these metrics with the data being examined in order to get a full picture on what is happening.