

To build an ML model in AWS and expose it through an API, the following steps can be followed:

- 1. Build and train the model in Amazon SageMaker:** This is an AWS service that allows from data collection and preparation, to training and deploying ML models quickly and easily. In addition, you can choose the type of compute instance to fit your budget and level of demand.
- 2. Create an Endpoint for the model:** Once the model has been trained in SageMaker, an Endpoint must be created to expose it through an API. The Endpoint is the URL of the model in SageMaker that will be used to perform the inferences.
- 3. Configure Amazon API Gateway:** This is a service that allows you to create, publish, monitor and protect APIs in a simple and scalable way. An API must be created that references the Endpoint of the model in SageMaker.
- 4. Configure AWS Lambda:** This is a serverless computing service that automatically executes code in response to events, such as API requests in API Gateway. An AWS Lambda function must be created to act as an intermediary between API Gateway and the model Endpoint in SageMaker. The Lambda function will receive the API requests, process them and send them to the Model Endpoint in SageMaker to get the response.
- 5. Set permissions and security:** It is important to ensure that only authorized users can access the model and make inferences. IAM (Identity and Access Management) can be used to manage permissions and API authentication and authorization.
- 6. Testing and deployment:** Once all the components are configured, tests should be performed to ensure that the model works correctly and that the API responds properly to requests. If everything works well, the model and API can be deployed in production