Introduction to Model Deployment

Agenda

- Introduction to Model Deployment
- Model Serialization
- API and API Endpoint
- HTTP Methods
- Handling Dependencies
- Handling Security

Let's begin the discussion by answering a few questions on Model Deployment

Model Deployment and Serialization

A financial risk assessment model that was serialized and deployed in production started throwing errors. Investigation revealed that one of the categorical feature contained new values not seen during training. What is the most effective strategy to prevent this issue in the future?

A. Reduce the learning rate during training to improve model generalization

B. Implement a fallback mechanism to assign unknown categories to an “Unknown” or "Other" class

C. Normalize all categorical values before feeding them into the model.

D. Increase the number of epochs during training to improve accuracy.

Model Deployment and Serialization

A financial risk assessment model that was serialized and deployed in production started throwing errors. Investigation revealed that one of the categorical feature contained new values not seen during training. What is the most effective strategy to prevent this issue in the future?

A. Reduce the learning rate during training to improve model generalization

B. Implement a fallback mechanism to assign unknown categories to an “Unknown” or "Other" class

C. Normalize all categorical values before feeding them into the model.

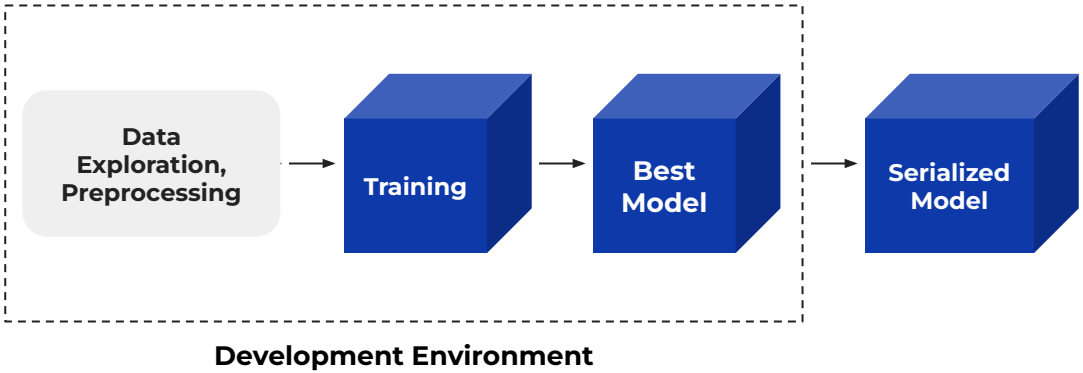
D. Increase the number of epochs during training to improve accuracy.

Model Deployment and Serialization

Deploying a model means taking a trained ML model, **packaging it**, and **setting it up for inference**.

Inference is the process of **inputting data points** into a machine learning model to **generate an output**.

Serialization is the process of translating a data structure or object state into a format that can be stored or transmitted and reconstructed later.

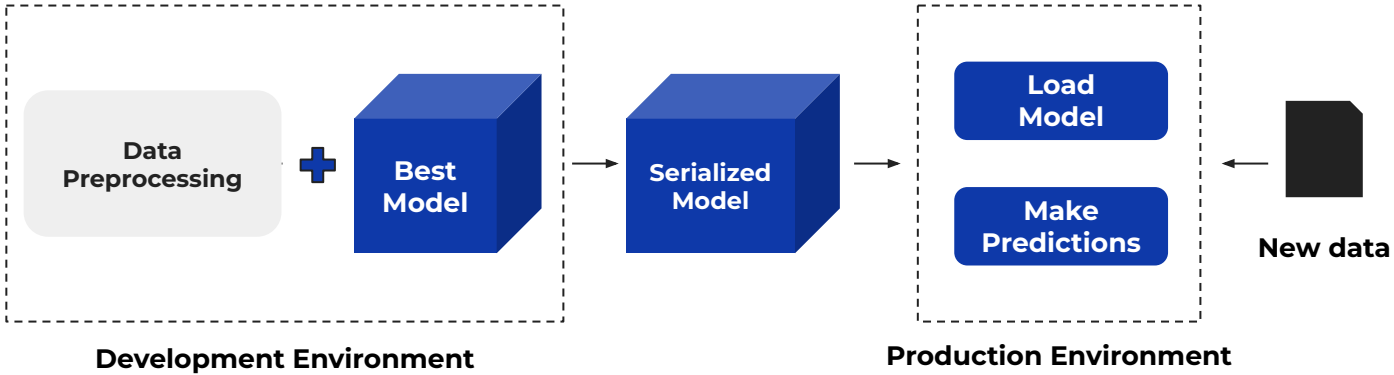


Model Deployment and Serialization

If new data had additional values in a categorical variables beyond what was used to train the model, it'll raise errors.

We need to **serialize the best model** along with the **data preprocessing steps** that were used to preprocess the data used to train the model.

We can build the data preprocessing pipeline such that new categories observed in new data in production are treated as a new category (like "Unknown" or "Other")



API and HTTP Status Codes

We developed an API to respond with the likelihood of customer churn based on customer attributes. A user makes a request to a resource that does not exist.
Which HTTP status code should the API return to indicate that the resource was not found?

A 401

C 500

B 200

D 404

API and HTTP Status Codes

We developed an API to respond with the likelihood of customer churn based on customer attributes. A user makes a request to a resource that does not exist.
Which HTTP status code should the API return to indicate that the resource was not found?

A. 401

C. 500

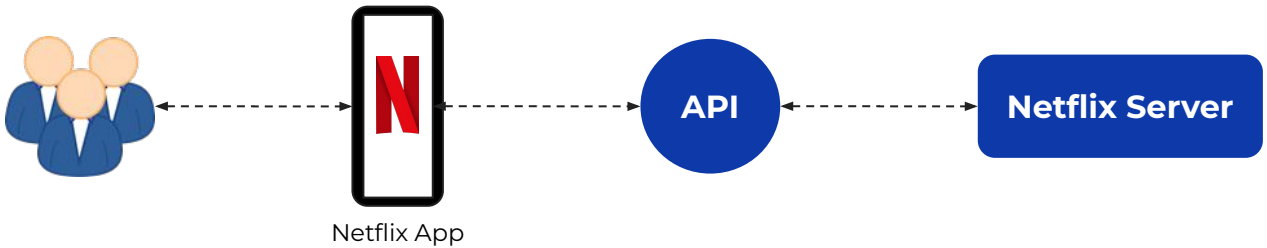
B. 200

D. 404

API and HTTP Status Codes

Application **P**rogramming **I**nterfaces, or **API**s, are **mechanisms** that **enable two software components to communicate** with each other **using a set of definitions and protocols**.

An **API request**, or an **API call**, is a **message sent to a server** asking an API to **provide a service or information**.



API and HTTP Status Codes

Once an API request is processed, it is important for the API to respond with a status indicating whether it succeeded or failed.

HTTP status codes provide this information to ensure smooth communication

Status Code	Status Message	Description
200	Ok	Request was successful
400	Bad Request	The request was invalid
401	Unauthorised	Authentication is required
403	Forbidden	Access is not allowed
404	Not Found	Requested resource does not exist
500	Internal Server Error	The server encountered an error

Endpoints

We created an API and want to enable different endpoints for various functionalities.
What is the purpose of defining separate endpoints?

A. To provide distinct functionalities, improving clarity and manageability

B. To slow down performance for better security

C. To enforce user limits on resource access

D. To reduce server load to an absolute minimum

Endpoints

We created an API and want to enable different endpoints for various functionalities.
What is the purpose of defining separate endpoints?

A. To provide distinct functionalities, improving clarity and manageability

B. To slow down performance for better security

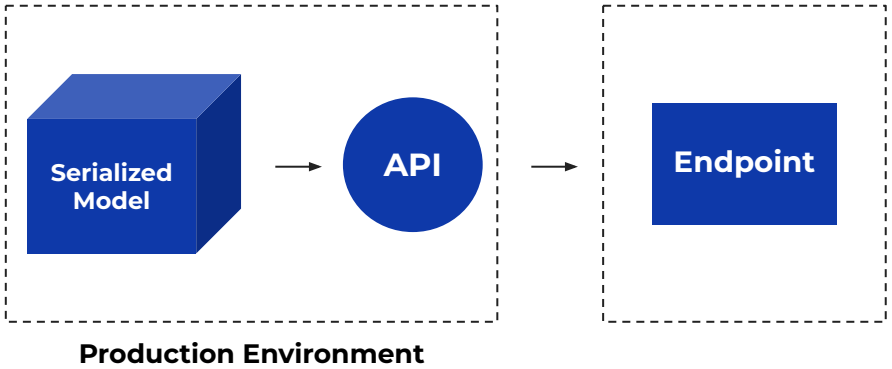
C. To enforce user limits on resource access

D. To reduce server load to an absolute minimum

Endpoints

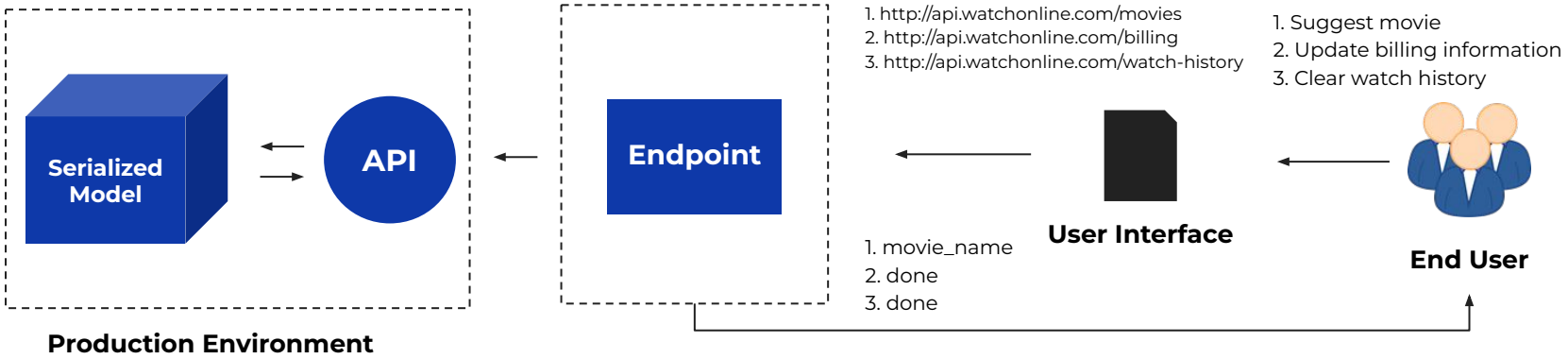
We can **create an endpoint from an API** to make things **simpler for the end user**

An **API endpoint** is a **digital location** where an **API receives API requests**, for resources on its server.



Endpoints

Each functionality (API request) is performed via an endpoint (URL), which is generally is a user interface



By structuring APIs with separate endpoints, developers can **enhance clarity and maintainability** by organizing different functionalities into distinct URL paths.

This approach ensures that each endpoint serves a specific purpose, making it easier to scale, debug, and optimize the API without affecting unrelated functionalities.

Role of Dependencies

A deployed ML model works on the developer's local system but fails on the cloud server due to a `ModuleNotFoundError`.
What is the best way to prevent such dependency issues in the future?

A. Manually install each package on the server

B. Use a `requirements.txt` file to define and install all dependencies

C. Use `pip install` without specifying versions to always get the latest libraries

D. Reinstall the entire Python environment each time before deployment

Role of Dependencies

A deployed ML model works on the developer's local system but fails on the cloud server due to a `ModuleNotFoundError`.
What is the best way to prevent such dependency issues in the future?

A. Manually install each package on the server

B. Use a `requirements.txt` file to define and install all dependencies

C. Use `pip install` without specifying versions to always get the latest libraries

D. Reinstall the entire Python environment each time before deployment

Role of Dependencies

Dependency issues arise when the production environment lacks required libraries that were available in the development environment where the model was trained.

We can **create a dependencies file** that'll act as a **blueprint for recreating** the exact **environment** where the model was trained and tested (the development environment).

In general, dependencies files are specified as a text file (***`requirements.txt`***)

Creates a **stable and reproducible environment** , allowing the model to be deployed seamlessly and function as expected in production environment

`requirements.txt`

```
sklearn==1.6.1
pandas==2.2.3
numpy==2.2.4
..
..
```

Access Keys

To safeguard sensitive user information while accessing your model's API, you decide to implement access controls using access keys. Which statement best describes why this is important?

A. It slows down the user experience

B. It prevents unauthorized access and protects sensitive data

C. It reduces the amount of input data

D. It complicates the API unnecessarily

Access Keys

To safeguard sensitive user information while accessing your model's API, you decide to implement access controls using access keys. Which statement best describes why this is important?

A. It slows down the user experience

B. It prevents unauthorized access and protects sensitive data

C. It reduces the amount of input data

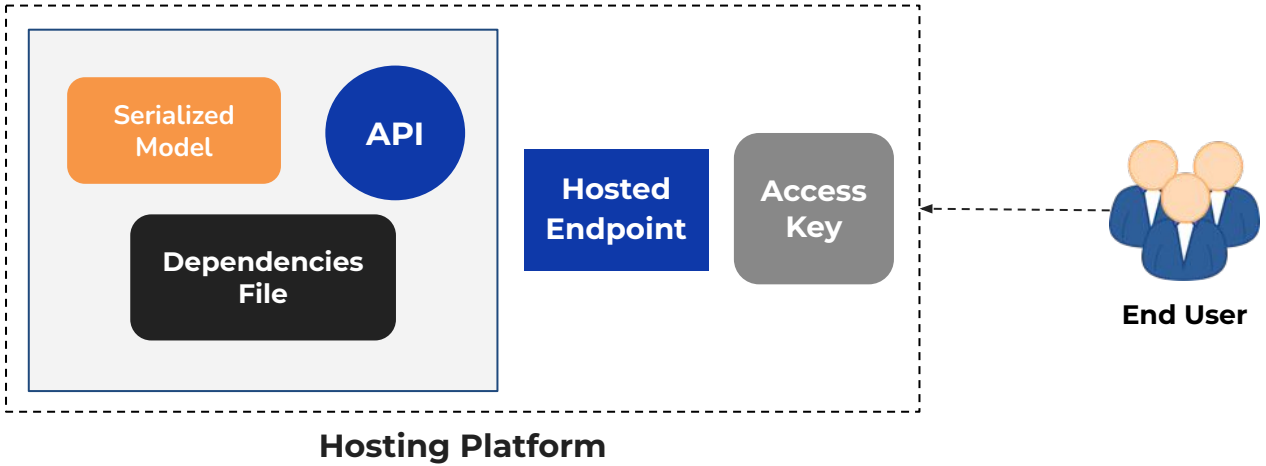
D. It complicates the API unnecessarily

Access Keys

An **access key** is a **unique identifier used to authenticate and authorize** a user or application to **access a secure resource**

Access keys prevent unauthorized access to secure resources (like data, models, and APIs)

Only authorized users can use the resources and functionality associated with a specific account



Great Learning

Happy Learning !