“Genre Classification and Recommendation System for Music on AWS Cloud Platform”

PROJECT REVIEW 1 REPORT

Submitted for the course: Cloud Computing (ITE3007)

By

NAME	REG. NO
AKSHAY RAGHAVAN	16BIT0070
ASTHA BARANWAL	16BIT0184

Slot D2+TD2

Name of Faculty: PROF. Priya V

SCHOOL OF INFORMATION TECHNOLOGY AND ENGINEERING

VIT UNIVERSITY

VIT®

UNIVERSITY

(Estd. u/s 3 of UGC Act 1956)

VELLORE ■ CHENNAI

www.vit.ac.in

ABSTRACT

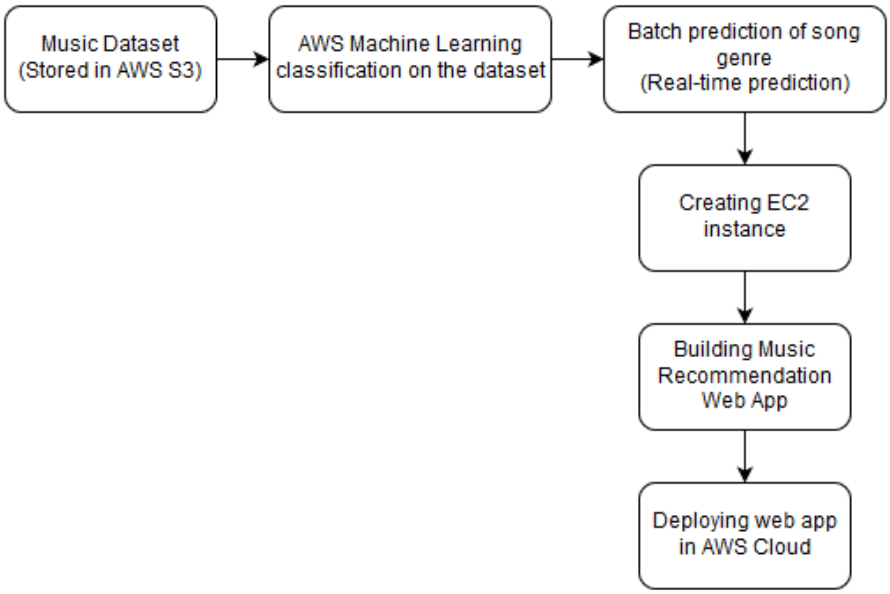
Music genre classification means segregating songs into genres like, Pop, Country, Rock, Jazz, R&B, Classical, etc. Genre classification is a very subjective topic. Based on the approach, songs might get classified differently by different people. It is fairly simple for a human being to identify the genre of a song. One thinks about how fast the beat of the song is, the mood of the song and the ambiance it creates. All these help create a mental picture of the song and thus the genres associated with it are determined.

That makes music genre classification a very controversial topic. Hence, creating an automated system based on a scientific approach is beneficial to the ever-growing music industry. Genre classification is crucial to the music industry as it plays a vital role in song recommendation systems. Song recommendation systems work in a way that help listeners find similar music.

Amazon Web Service (AWS) is a cloud platform that helps create robust and accurate machine learning models without any coding. It allows deployment of apps on cloud platform. This project uses AWS services to deploy a song recommendation web app.

Songs will be classified into genres using the machine learning service of AWS. The dataset we will be using is a subset of the Million Song dataset from the CORGIS dataset project. After a machine learning classification model is built, genre of batches of songs can be predicted using the real-time prediction API created. These songs will then be used in the Music Recommendation web app hosted on AWS cloud platform via an EC2 instance to recommend similar music of the same genre to the users.

PROJECT FLOWCHART



PLATFORMS AND LANGUAGES

The project will be hosted entirely on the AWS Cloud Platform:

AWS Machine Learning

Amazon Machine Learning is a managed service for building ML models and generating predictions, enabling the development of robust, scalable smart applications. Amazon Machine Learning enables you to use powerful machine learning technology without requiring an extensive background in machine learning algorithms and techniques.

The Music Genre Classification System will be built on the Machine learning service of AWS.

AWS EC2

Amazon Elastic Compute Cloud (EC2) forms a central part of AWS by allowing users to rent virtual computers on which to run their own computer applications. EC2 encourages scalable deployment of applications by providing a web service through which a user can boot an Amazon Machine Image (AMI) to configure a virtual machine, which Amazon calls an "instance", containing any software desired. A user can create, launch, and terminate server-instances as needed.

EC2 will be used to host the Music Recommendation System web app. XAMPP and Apache server will be installed within the EC2 instance created. The web app will be built in PHP and MySQL.

AWS S3

Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance.

The dataset for music genre classification will be stored in a S3 bucket.

PuTTy and PuTTygen

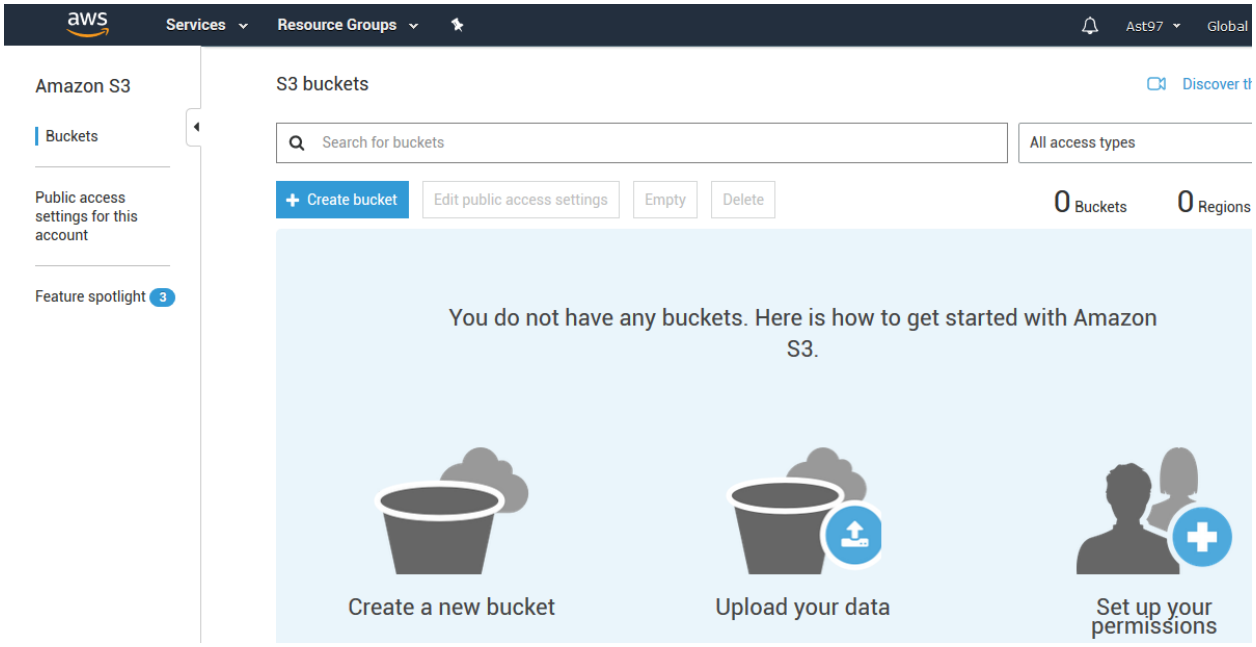
PuTTY is an SSH and telnet client. PuTTYgen is a key generator tool for creating SSH keys for PuTTY. They are used to connect to the EC2 instances created.

PROJECT WORKFLOW

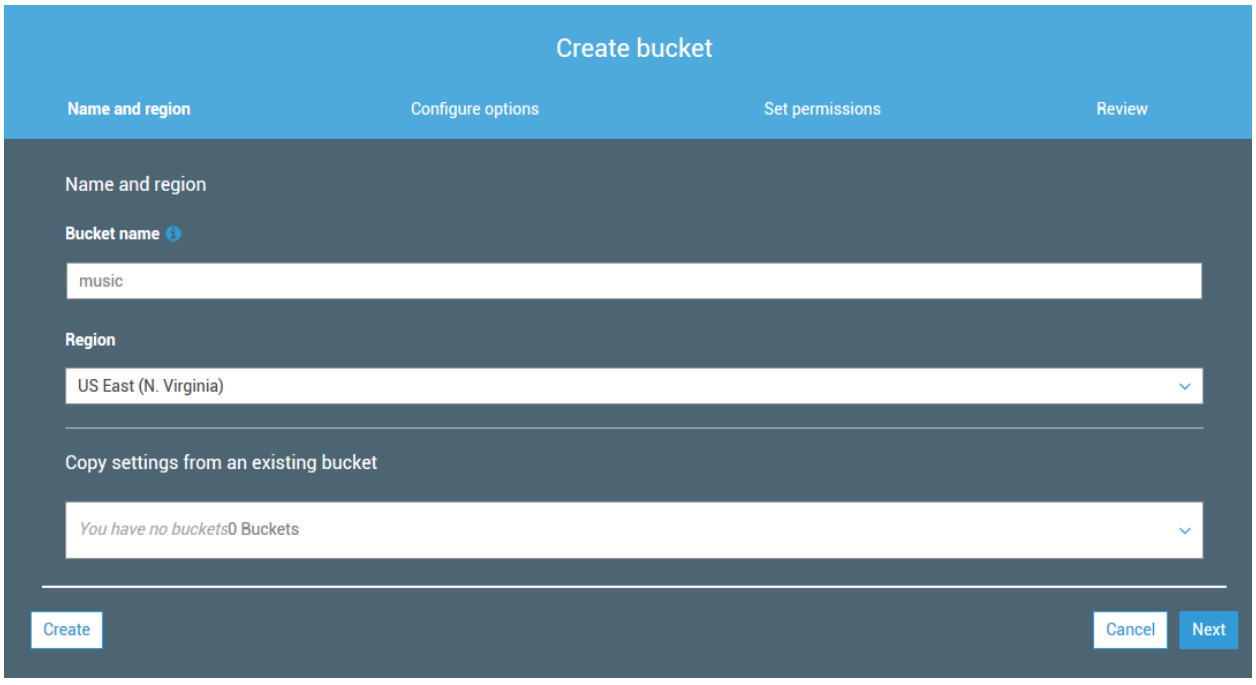
THE AWS ML MODEL

S3 BUCKET

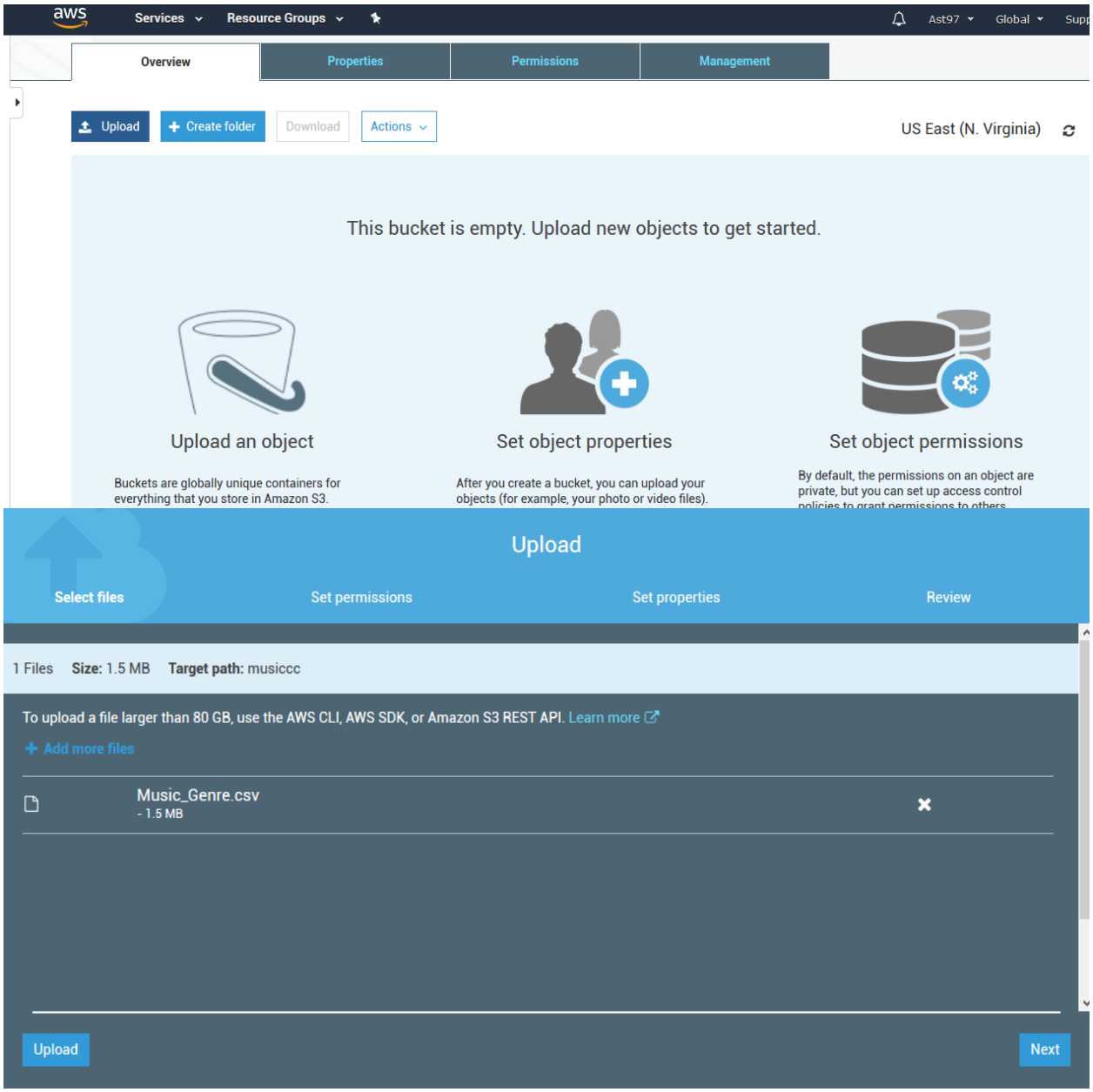
Go to S3 and click on the ‘Create bucket’ button.



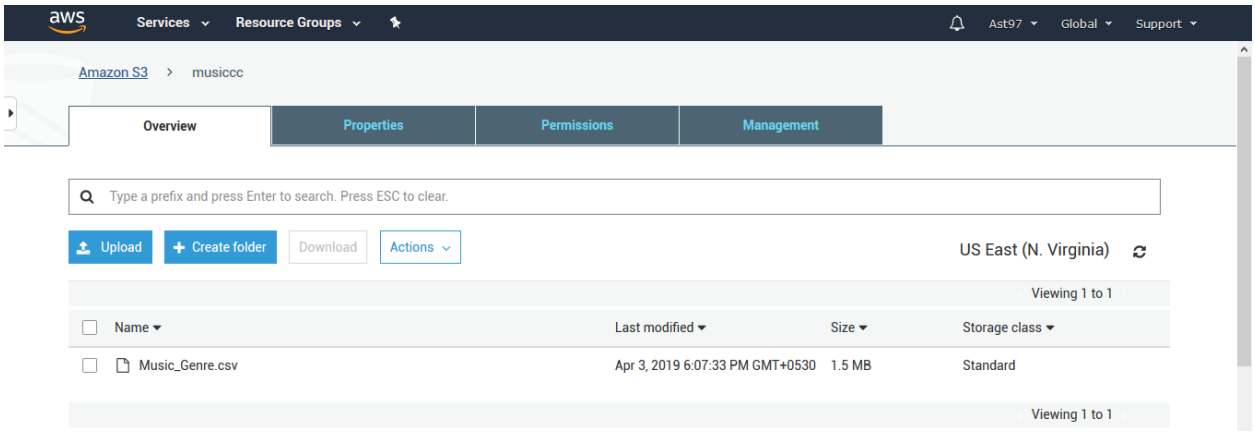
Enter the bucket name ‘musiccc’ and click on ‘Create’



Click on ‘Upload’ to upload the dataset to cloud

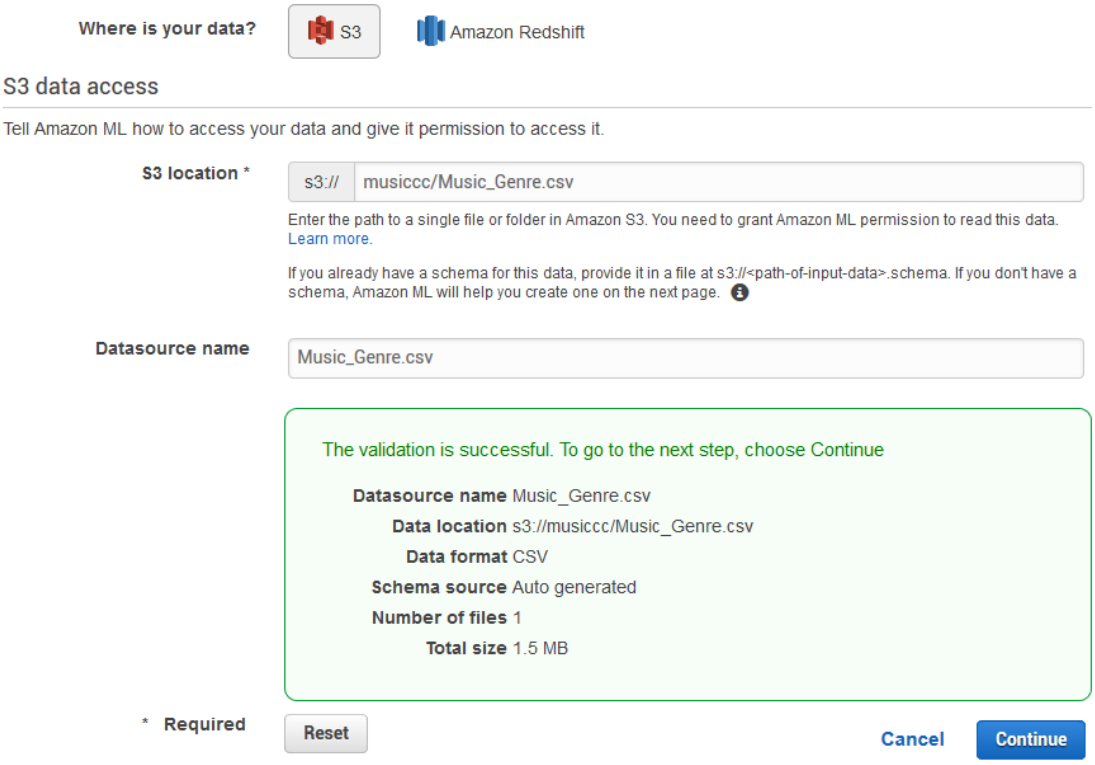


The file, ‘Music_Genre.csv’ is uploaded in the music bucket. Now this dataset can be used to build the AWS ML model for music genre classification.



AWS MACHINE LEARNING

Select the dataset from S3 Bucket 'musi



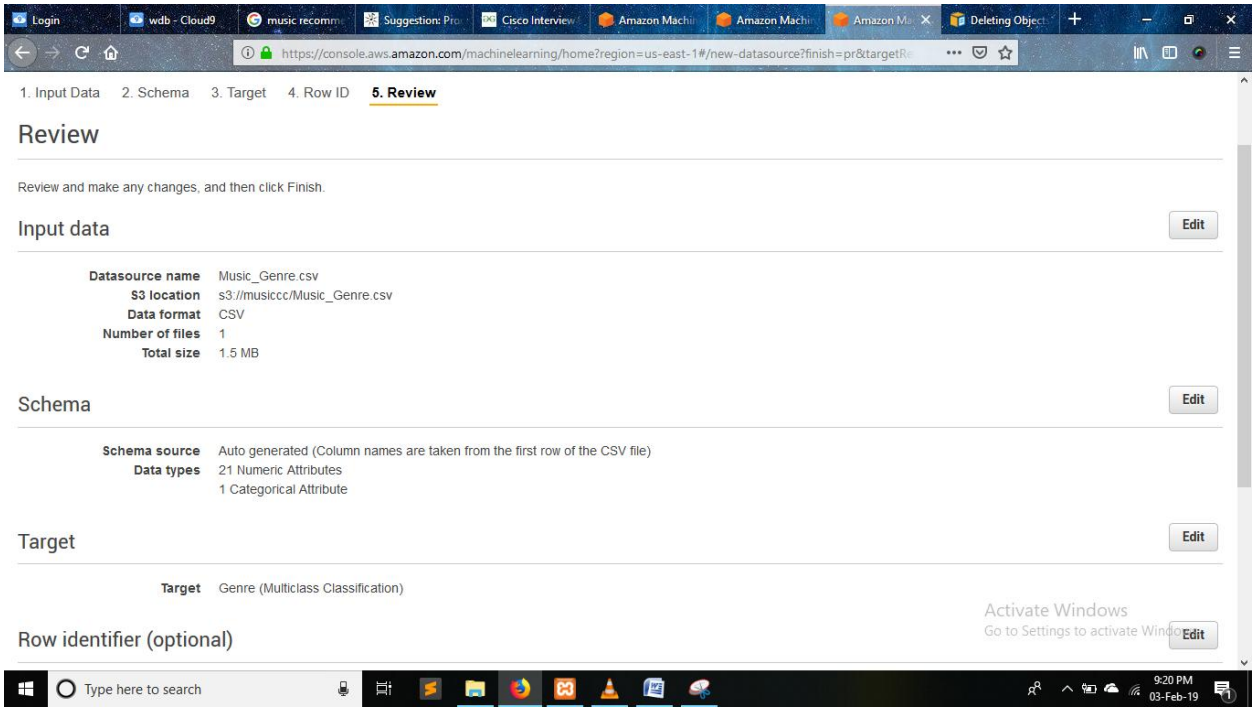
ccc'

ACTION: Change type

Search by attribute name

Items per page: 10 | 1 - 10 of 22

		Name	Data type	Sample field value 1	Sample field value 2	Sample field value 3
☐	1	Song_popularity	Numeric	0.187895	0.1904	0.190656
☐	2	Artist_popularity	Numeric	0.159717	0.31312	0.259463
☐	3	Duration	Numeric	232.202	153.33831999999998	200.25423999999998
☐	4	Bars_confidence	Numeric	0.37	0.185	0.031
☐	5	Bars_start	Numeric	0.37233	9.92414	1.64265
☐	6	Beats_confidence	Numeric	0.0	0.376	0.425
☐	7	Beats_start	Numeric	0.37233	0.1937	1.64265
☐	8	End_of_fade_in	Numeric	3.5580000000000003	0.166	0.0
☐	9	Familiarity	Numeric	0.39427399999999996	0.5052369999999999	0.433721
☐	10	Key	Numeric	2.0	8.0	7.0



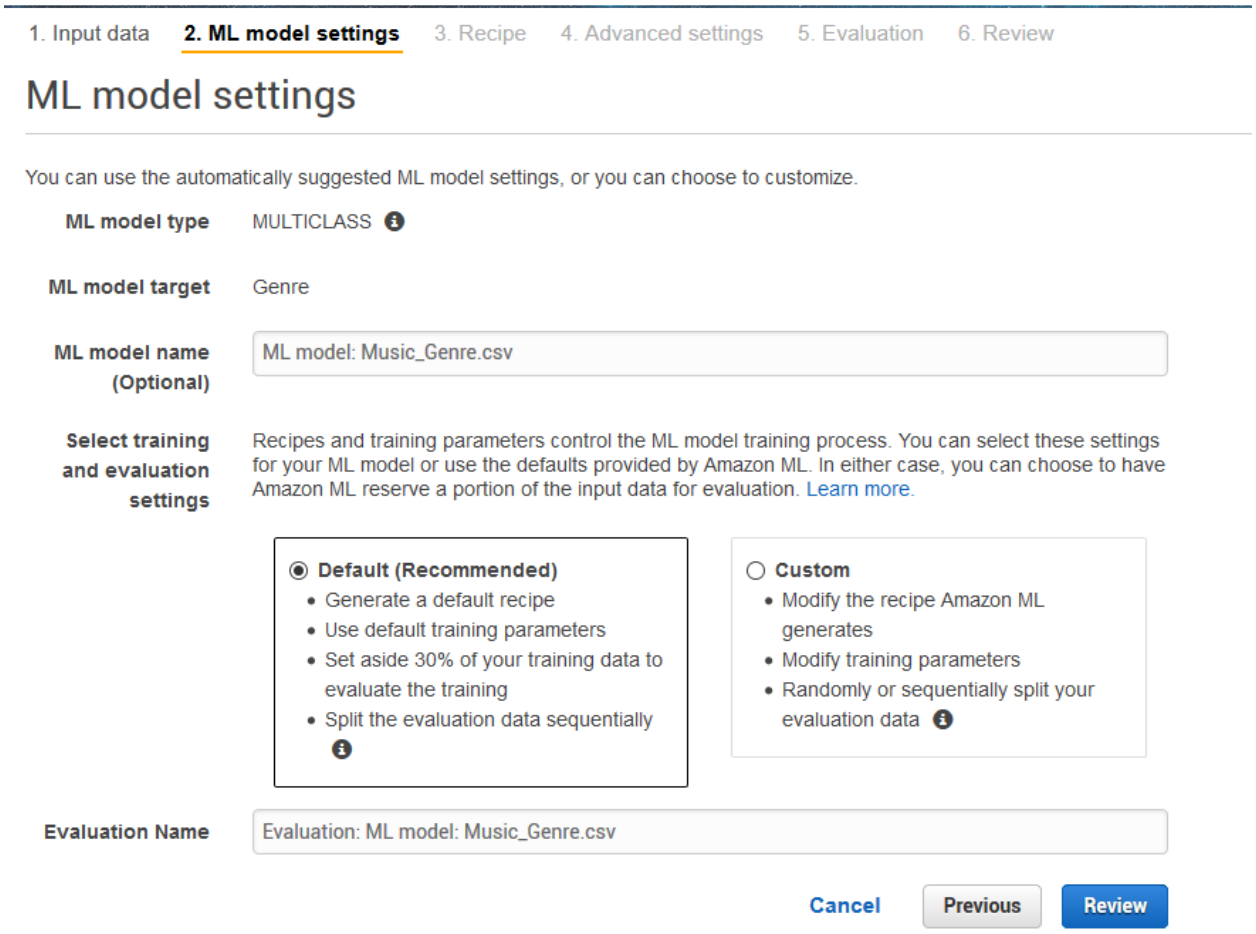
For Training and evaluation settings, ensure that Default is selected.

For Name this evaluation, accept the default, Evaluation: ML model: Music_Genre.

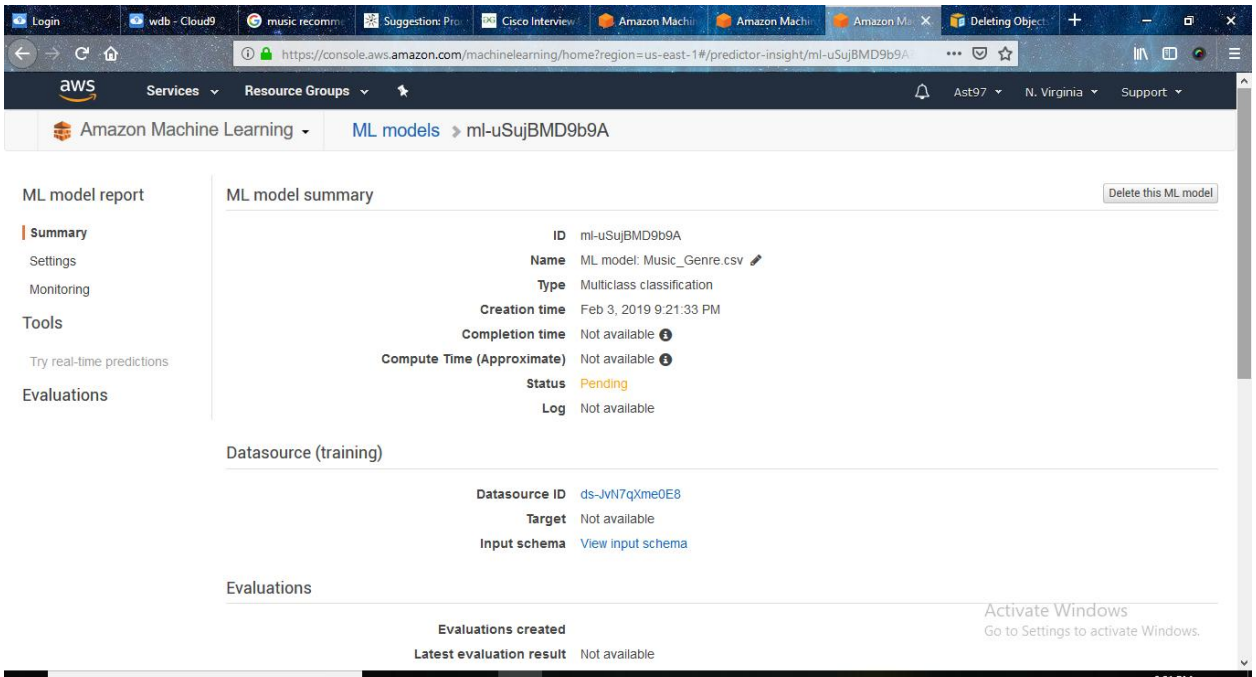
Choose Review, review your settings, and then choose Finish.

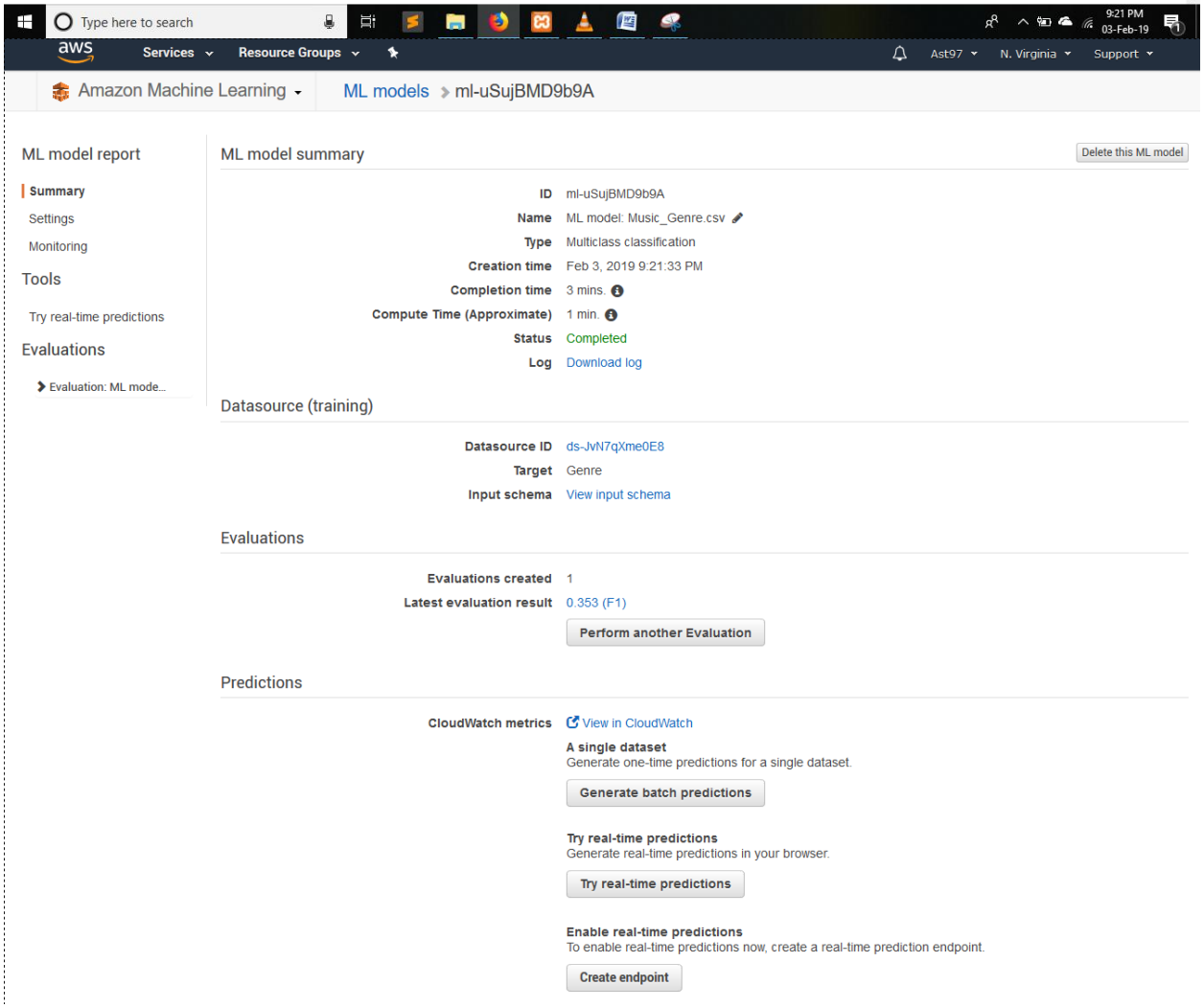
After you choose Finish, Amazon ML adds your model to the processing queue. When Amazon ML creates your model, it applies the defaults and performs the following actions:

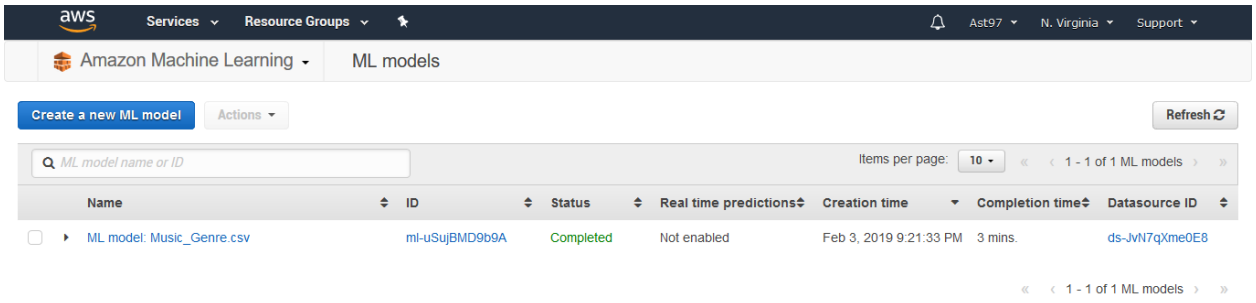
- Splits the training datasource into two sections, one containing 70% of the data and one containing the remaining 30%
- Trains the ML model on the section that contains 70% of the input data
- Evaluates the model using the remaining 30% of the input data



While your model is in the queue, Amazon ML reports the status as Pending. While Amazon ML creates your model, it reports the status as In Progress. When it has completed all actions, it reports the status as Completed. Wait for the evaluation to complete before proceeding.

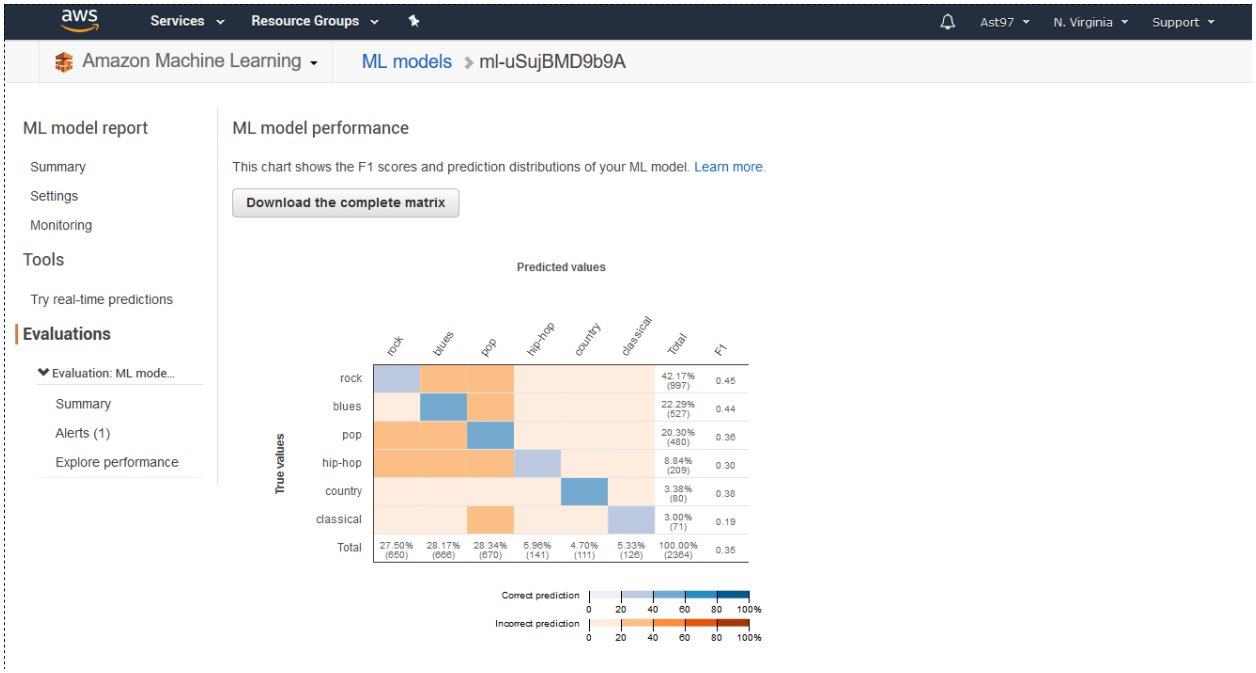

Login
wdb - Cloud9
music recomm
Suggestion: Pro
Cisco Interview
Amazon Machi
Amazon Machi
Amazon Ma
Deleting Object
https://console.aws.amazon.com/machinelearning/home?region=us-east-1#/predictor-insight/ml-uSujBMD9b9A
aws
Services
Resource Groups
Ast97
N. Virginia
Support
Amazon Machine Learning
ML models
ml-uSujBMD9b9A
ML model report
Summary
Settings
Monitoring
Tools
Try real-time predictions
Evaluations
ML model summary
Delete this ML model
ID ml-uSujBMD9b9A
Name ML model: Music_Genre.csv
Type Multiclass classification
Creation time Feb 3, 2019 9:21:33 PM
Completion time Not available
Compute Time (Approximate) Not available
Status Pending
Log Not available
Datasource (training)
Datasource ID ds-JvN7qXme0E8
Target Not available
Input schema View input schema
Evaluations
Evaluations created
Latest evaluation result Not available
Activate Windows
Go to Settings to activate Windows.

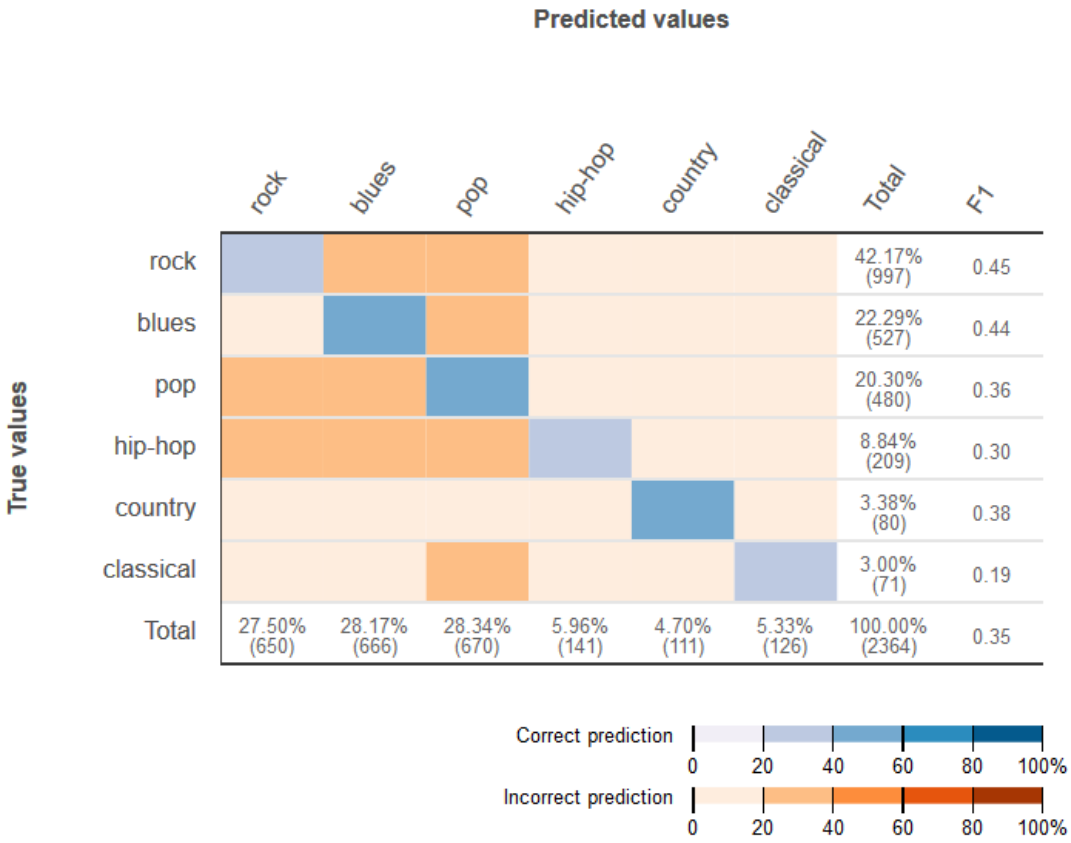

Type here to search
9:21 PM
03-Feb-19
aws
Services
Resource Groups
Ast97
N. Virginia
Support
Amazon Machine Learning
ML models
ml-uSujBMD9b9A
ML model report
Summary
Settings
Monitoring
Tools
Try real-time predictions
Evaluations
Evaluation: ML mode...
ML model summary
Delete this ML model
ID ml-uSujBMD9b9A
Name ML model: Music_Genre.csv
Type Multiclass classification
Creation time Feb 3, 2019 9:21:33 PM
Completion time 3 mins.
Compute Time (Approximate) 1 min.
Status Completed
Log Download log
Datasource (training)
Datasource ID ds-JvN7qXme0E8
Target Genre
Input schema View input schema
Evaluations
Evaluations created 1
Latest evaluation result 0.353 (F1)
Perform another Evaluation
Predictions
CloudWatch metrics View in CloudWatch
A single dataset
Generate one-time predictions for a single dataset.
Generate batch predictions
Try real-time predictions
Generate real-time predictions in your browser.
Try real-time predictions
Enable real-time predictions
To enable real-time predictions now, create a real-time prediction endpoint.
Create endpoint



On the ML model summary page, in the ML model report navigation pane, choose Evaluations, choose Evaluation: ML model: Music_Genre, and then choose Summary.

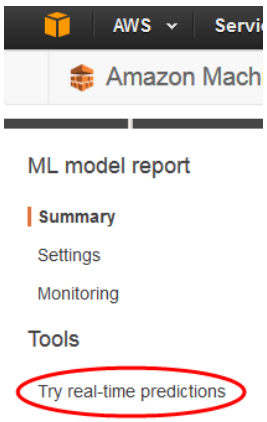
On the Evaluation summary page, review the evaluation summary, including the model's confusion matrix.





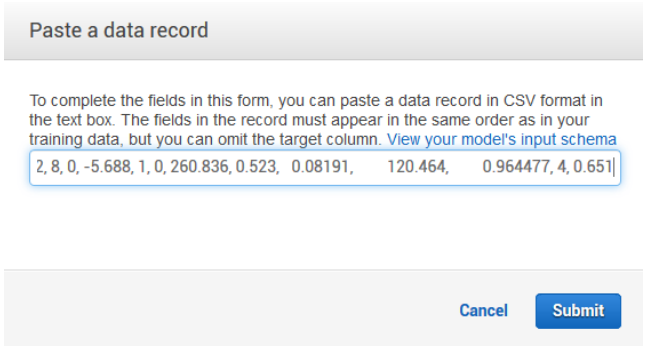
A real-time prediction is a prediction for a single observation that Amazon ML generates on demand.

In the ML model report navigation pane, choose Try real-time predictions.

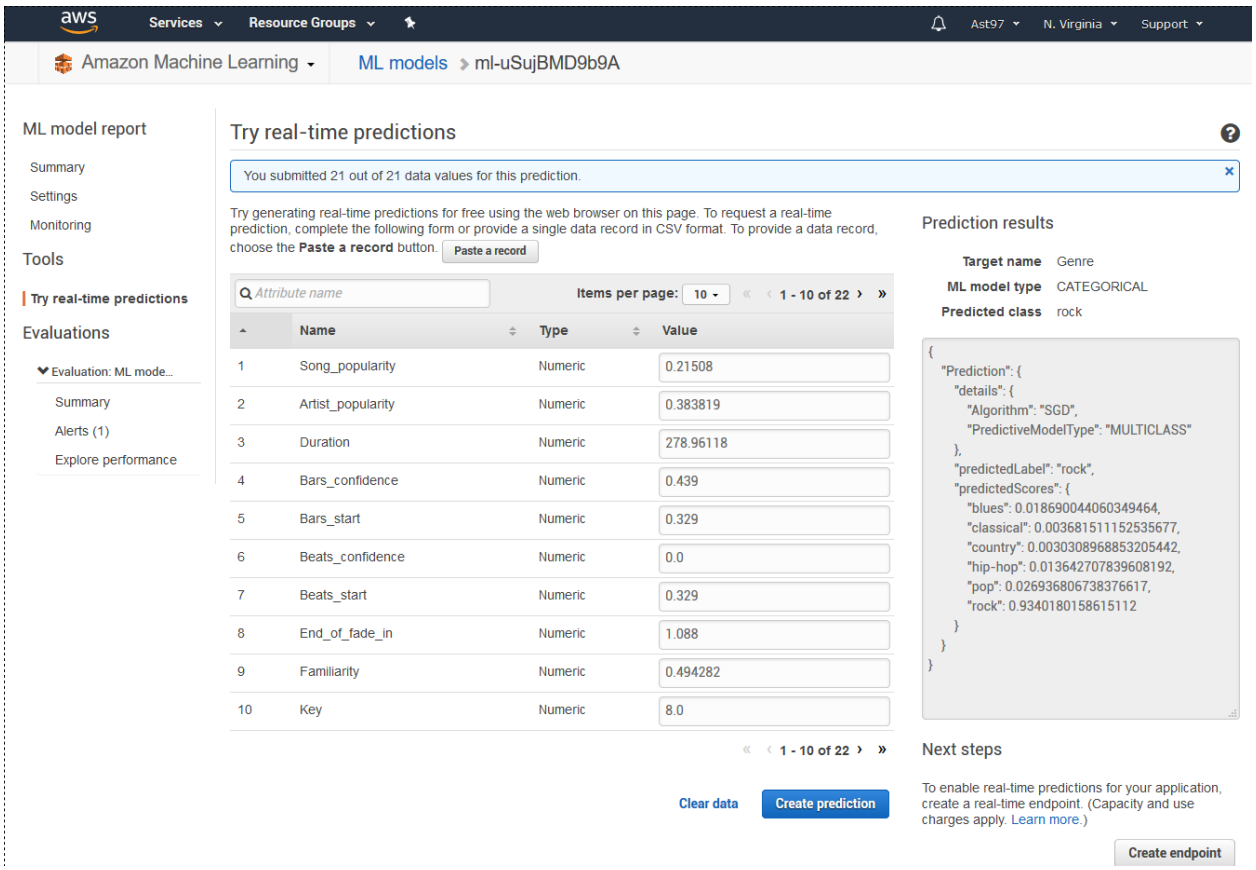


Choose Paste a record.

In the Paste a record dialog box, paste the following observation:



In the Paste a record dialog box, choose Submit to confirm that you want to generate a prediction for this observation. Amazon ML populates the values in the real-time prediction form. At the bottom of the page, choose Create prediction.



The prediction appears in the Prediction results pane on the right. This prediction has a Predicted label of the genre of the song which was entered for prediction.

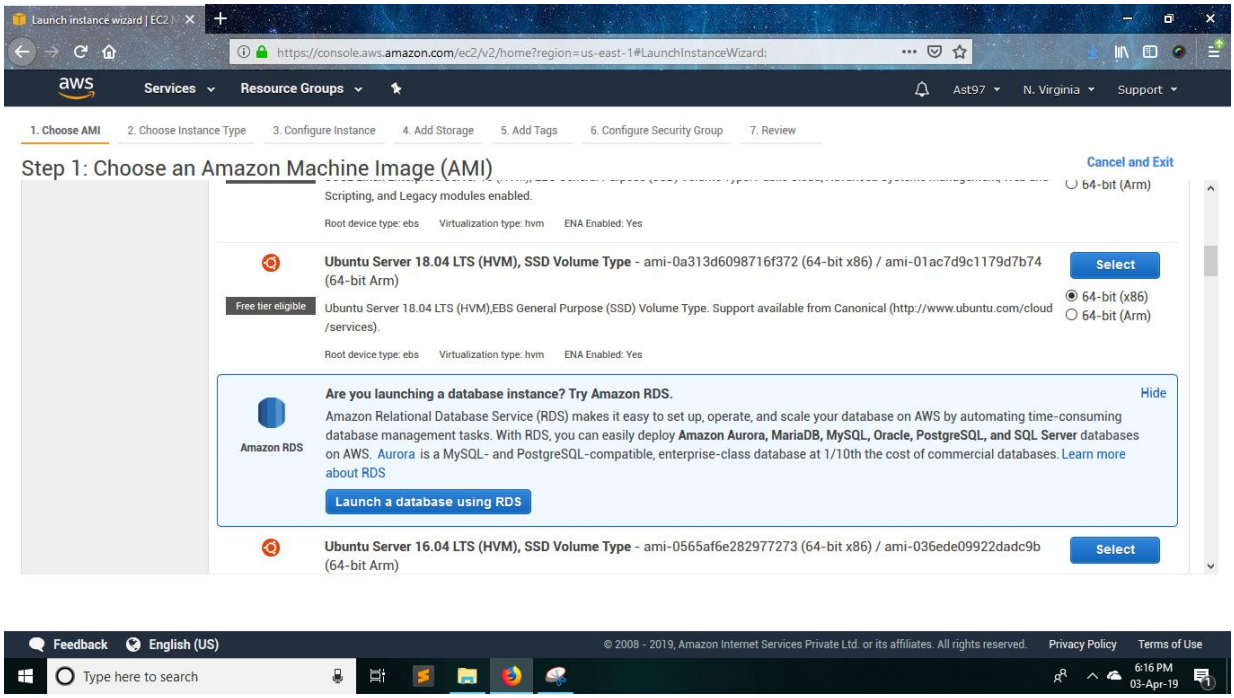
Prediction results

Target name	Genre
ML model type	CATEGORICAL
Predicted class	rock

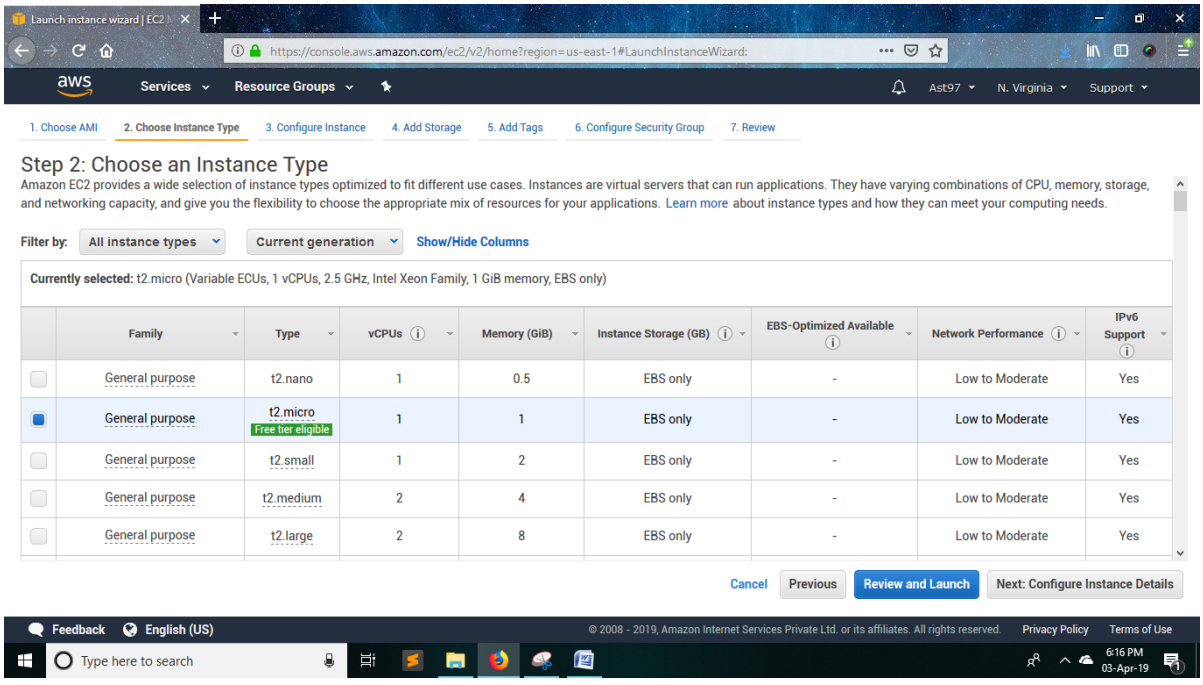

```
{
  "Prediction": {
    "details": {
      "Algorithm": "SGD",
      "PredictiveModelType": "MULTICLASS"
    },
    "predictedLabel": "rock",
    "predictedScores": {
      "blues": 0.018690044060349464,
      "classical": 0.003681511152535677,
      "country": 0.0030308968853205442,
      "hip-hop": 0.013642707839608192,
      "pop": 0.026936806738376617,
      "rock": 0.9340180158615112
    }
  }
}
```

THE SONG RECOMMENDATION SYSTEM

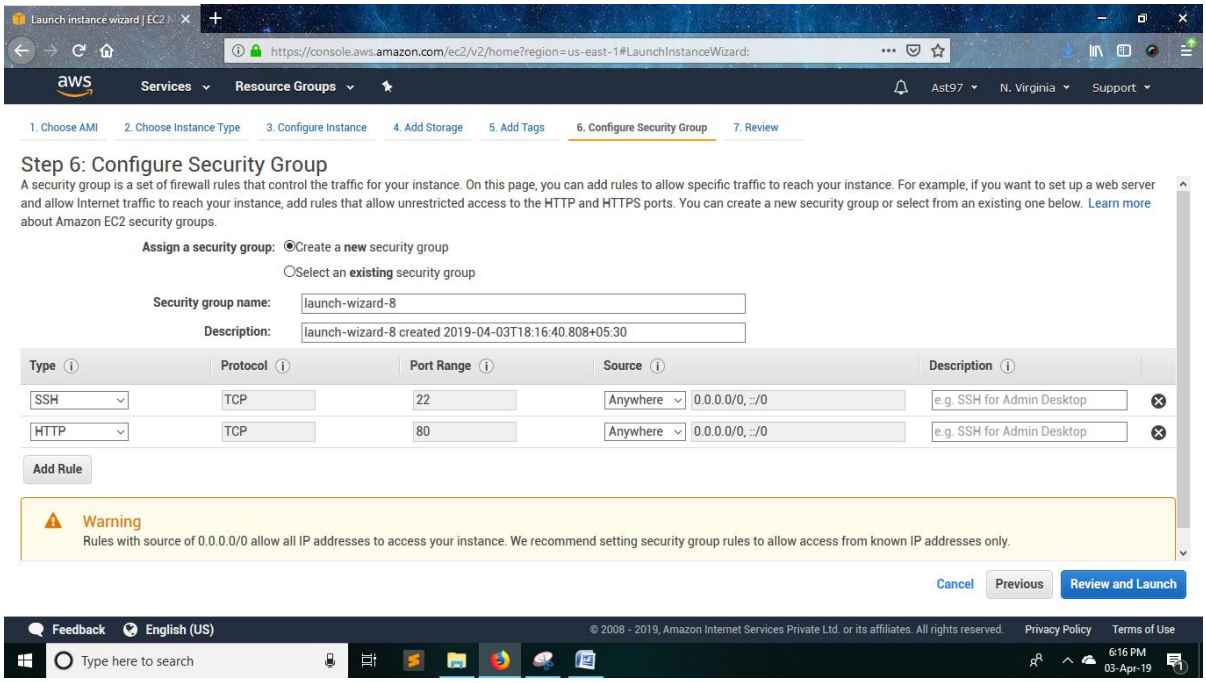
Select EC2 from Compute category. Click Launch Instance button. Now choose operating system for your web server, Ubuntu is suggested for more package advantages. Make sure, that should be listed in free tier.



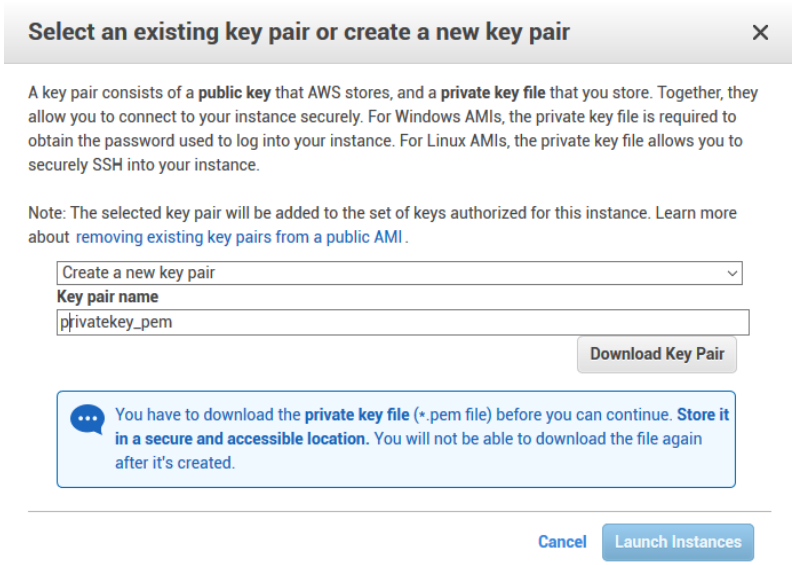
Here you can modify your server type. Free tier t2.micro is selected.

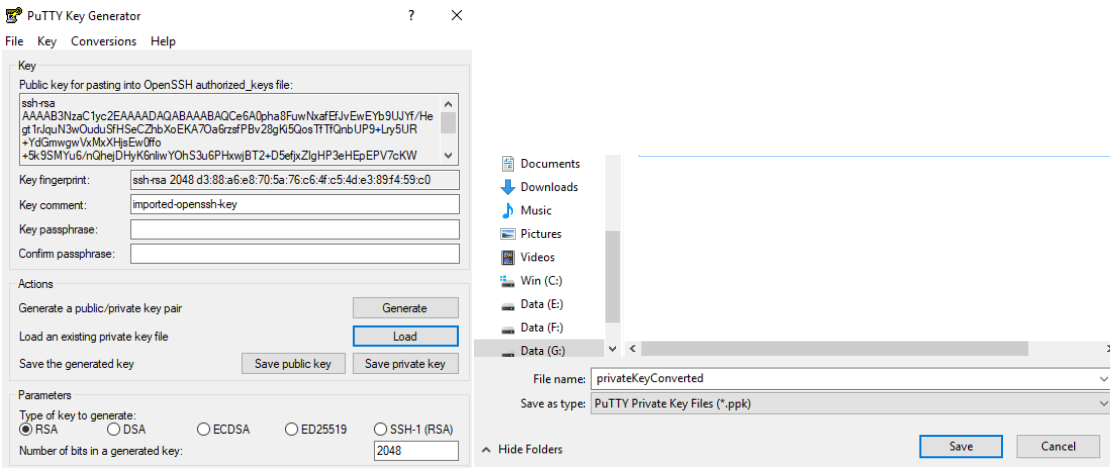


Security groups are modified. HTTP is added and source is set to ‘Anywhere’.

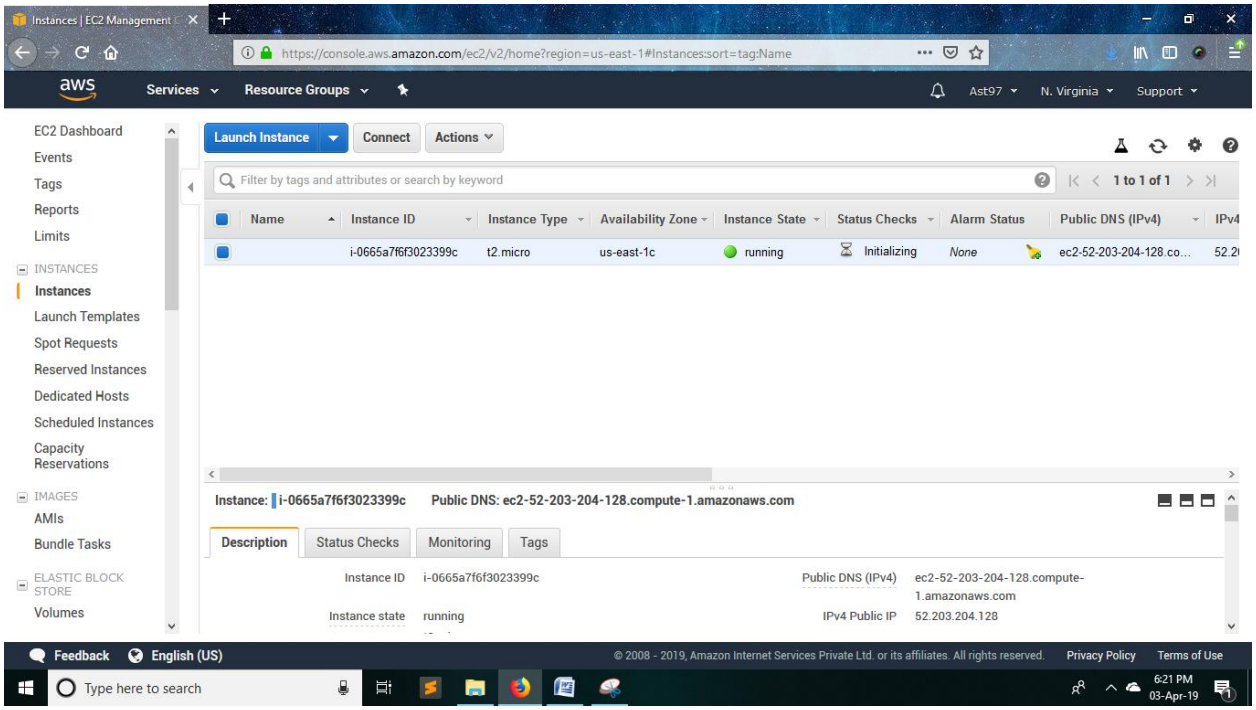


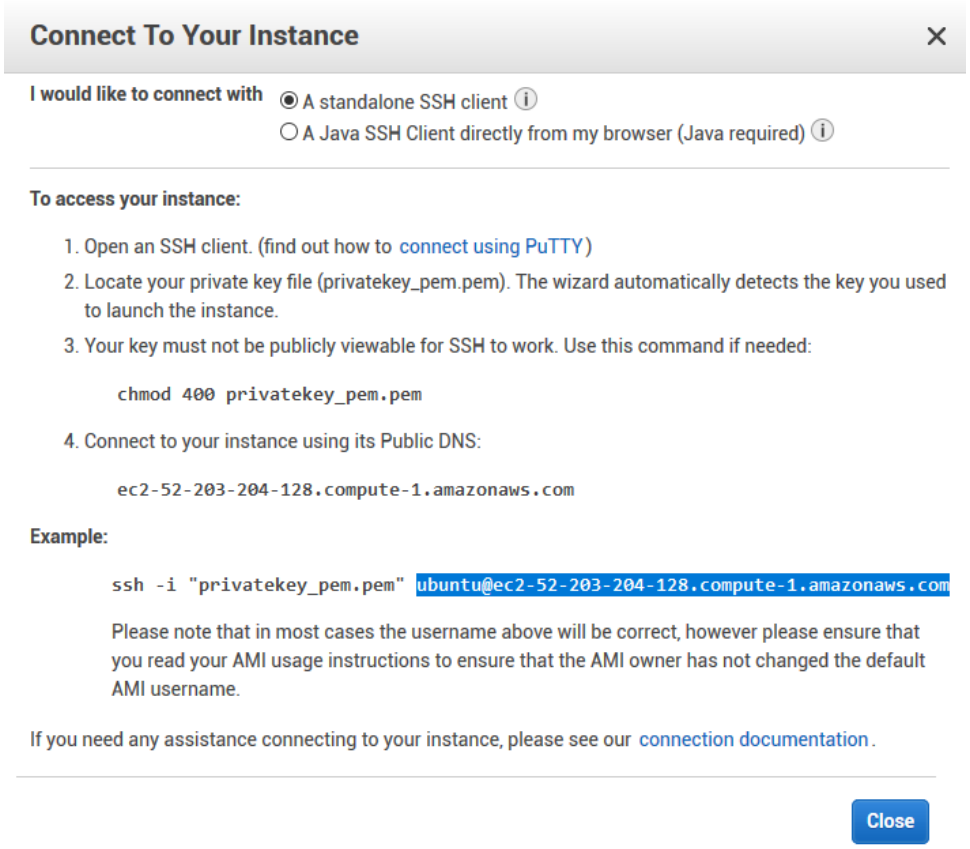
Here choose Create a new pair key. Give valid name and click to Download Key Pair. You will get an .pem file.



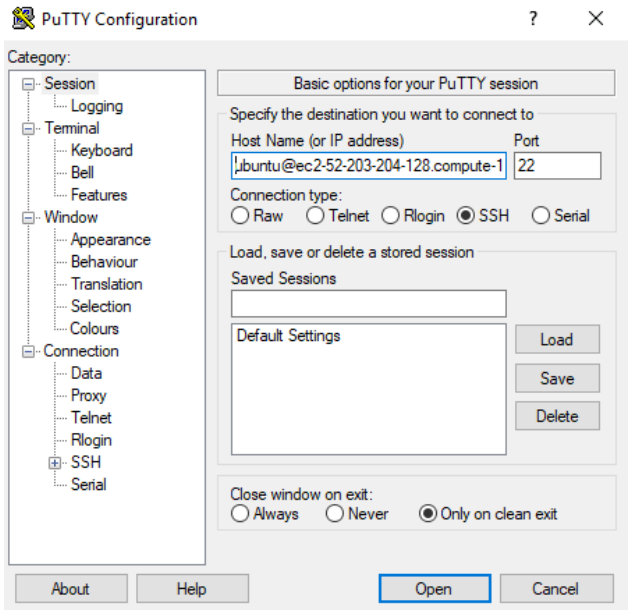


Instance has been created successfully.

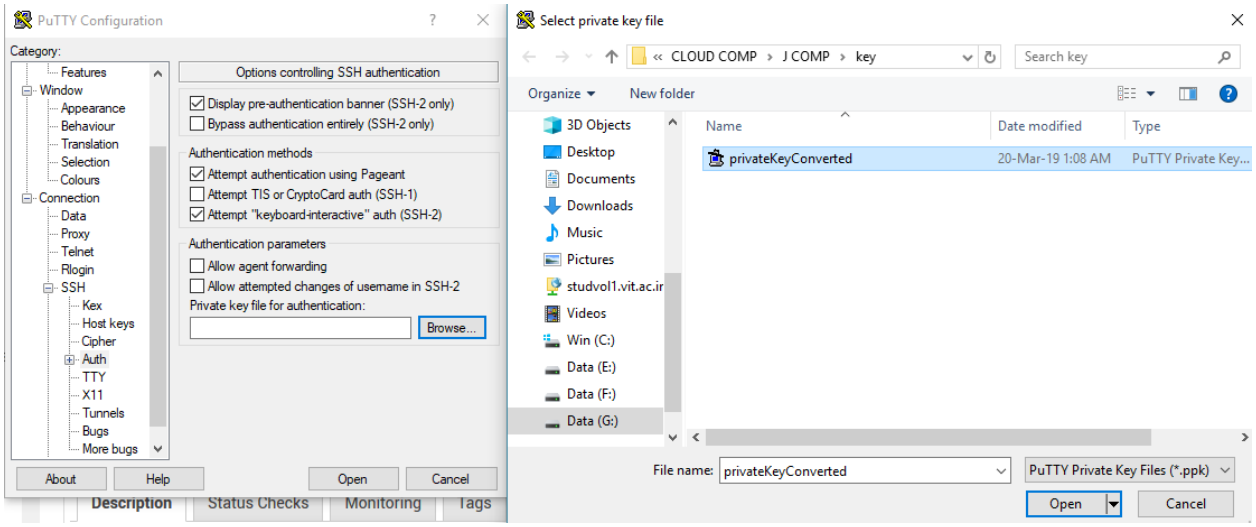




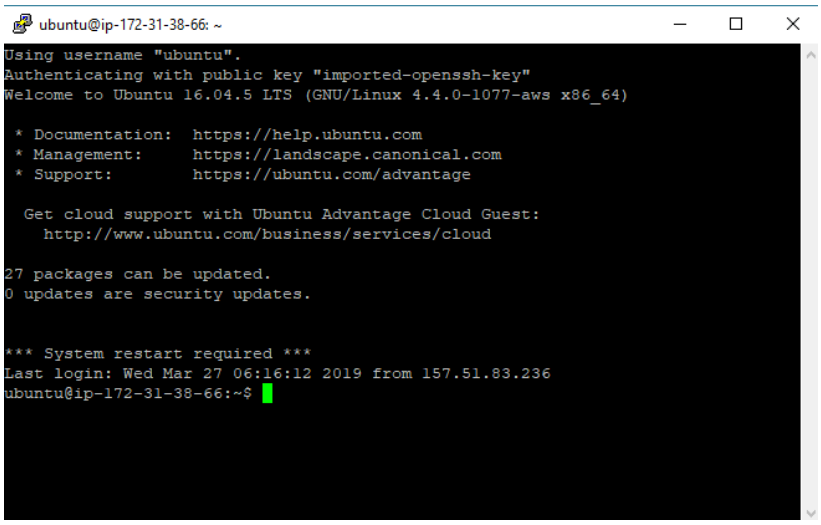
Copy and paste the IP from the ‘Connect To Your Instance’ dialog box



Go to SSH-> Auth and select the converted private key (.ppk format)



The connection has been established to the putty telnet client



Type in the following commands to install XAMPP:

Download XAMPP for 64 bit

```
wget https://www.apachefriends.org/xampp-files/7.0.23/xampp-linux-x64-7.0.23-0-installer.run
```

Make Execute Installation

```
sudo chmod +x xampp-linux-x64-7.0.23-0-installer.run
```

Run Installation

```
sudo ./xampp-linux-x64-7.0.23-0-installer.run
```

XAMPP instructions

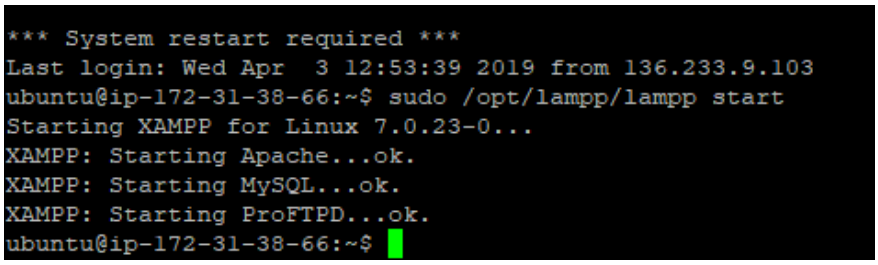
Select the components you want to install; clear the components you do not want to install. Click Next when you are ready to continue.

```
XAMPP Core Files : Y (Cannot be edited)
XAMPP Developer Files [Y/n] : Y
Is the selection above correct? [Y/n]: Y

Installation Directory
XAMPP will be installed to /opt/lampp
Press [Enter] to continue:
Do you want to continue? [Y/n]:Y
```

Run XAMPP

```
sudo /opt/lampp/lampp start
```



XAMPP Access Forbidden

Open your browser and access http://IP-ADDRESS/ you will find this Access forbidden screen.

XAMPP Configurations

Edit XAMPP configurations.

```
vi /opt/lampp/etc/extra/httpd-xampp.conf
```

```
<LocationMatch "^/(?i:(?:xampp|security|licenses|phpmyadmin|webalizer|server-status|server-info))">
Require local
ErrorDocument 403 /error/XAMPP_FORBIDDEN.html.var
</LocationMatch>
```

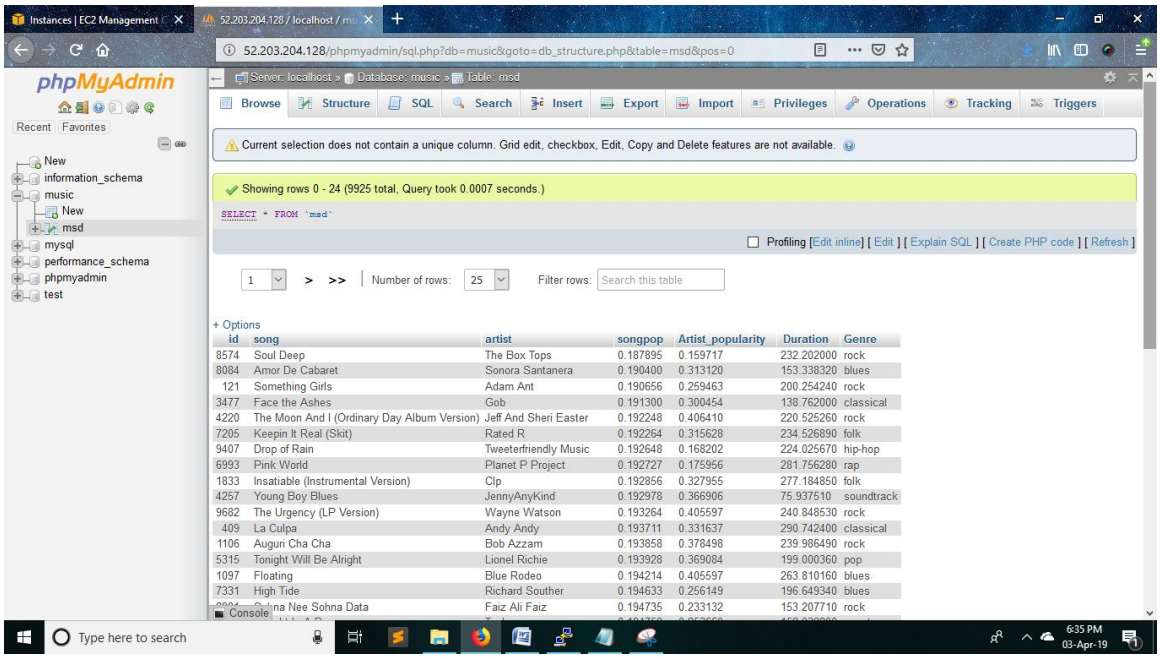
to

```
<LocationMatch "^/(?i:(?:xampp|security|licenses|phpmyadmin|webalizer|server-status|server-info))">
Order deny,allow
Allow from all
Allow from ::1 127.0.0.0/8 \
fc00::/7 10.0.0.0/8 172.16.0.0/12 192.168.0.0/16 \
fe80::/10 169.254.0.0/16
ErrorDocument 403 /error/XAMPP_FORBIDDEN.html.var
</LocationMatch>
```

Restart XAMPP

```
sudo /opt/lampp/lampp restart
```

Create a database ‘music’ on phpmyadmin and import the dataset as msd table

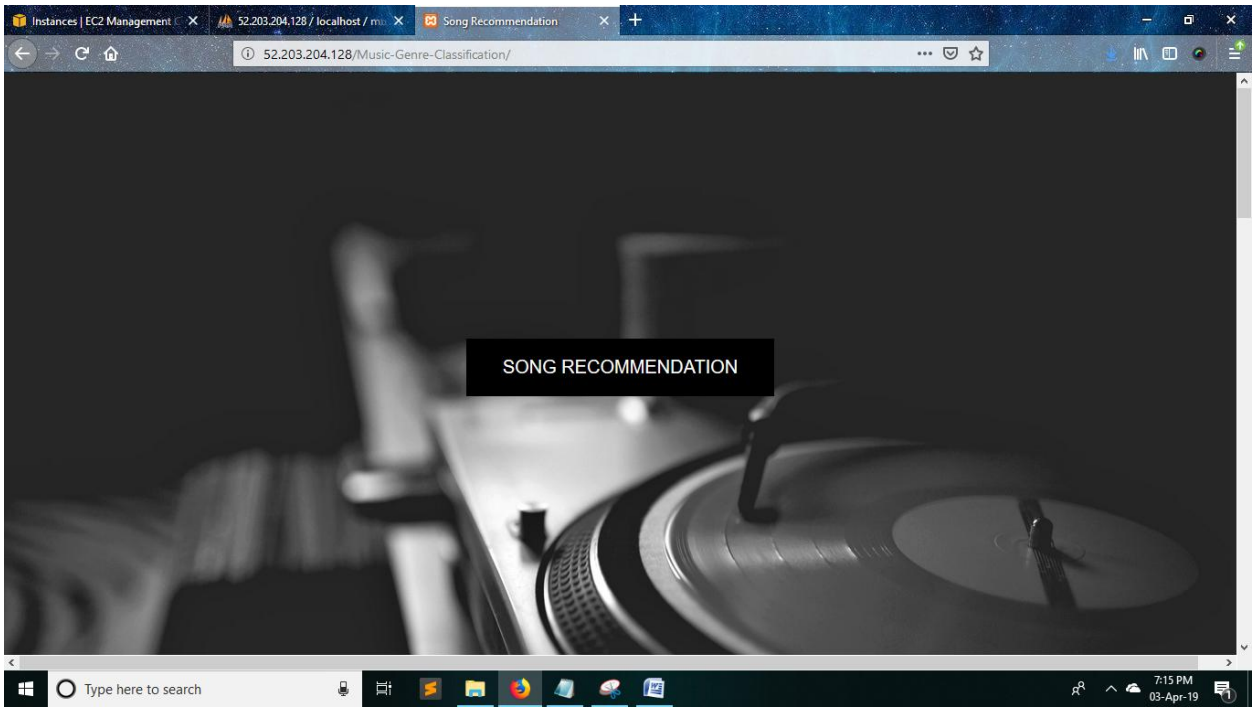


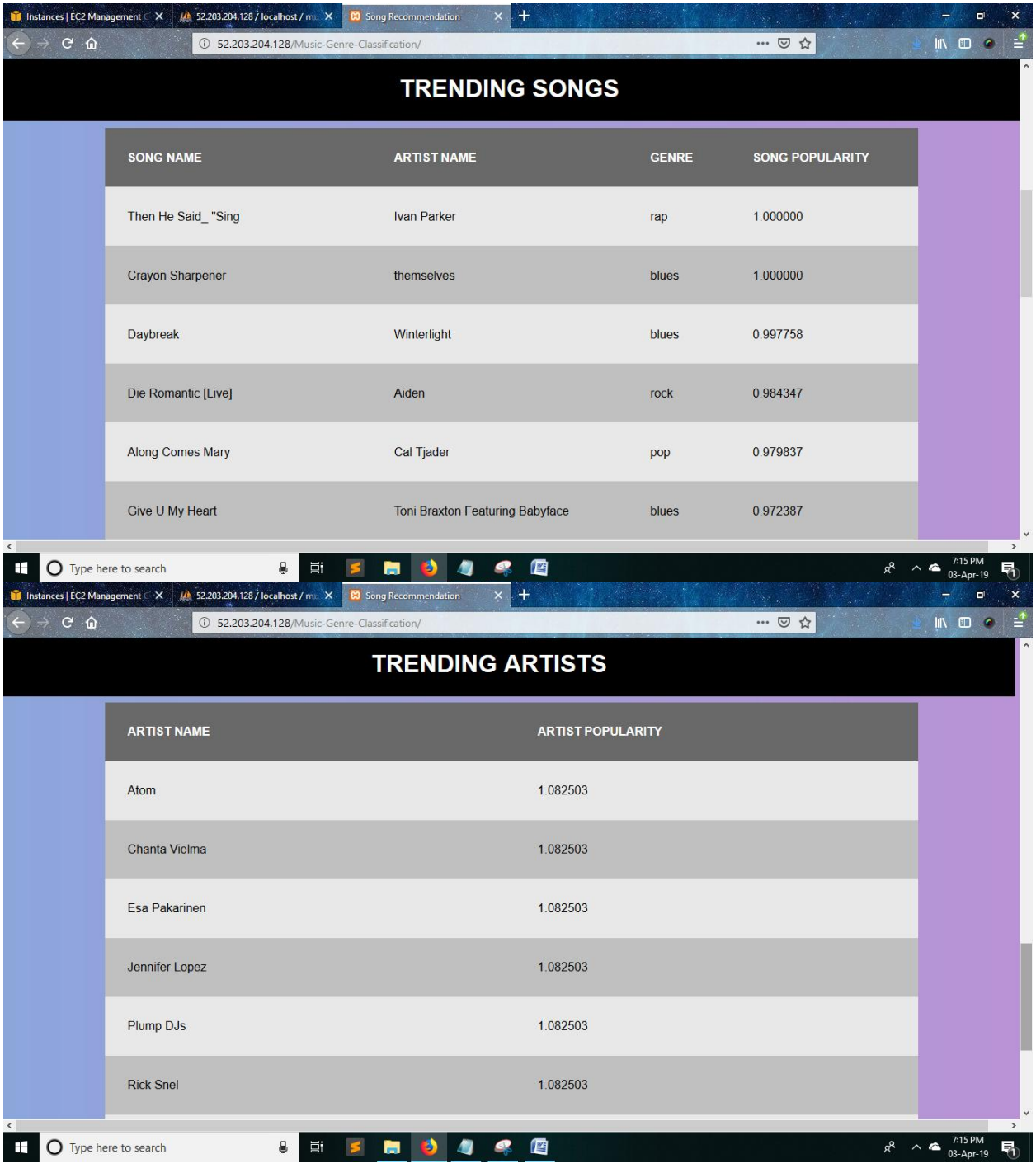

```
cd /opt/lampp/htdocs/
sudo chown -R $(whoami):$(whoami) /var/..
```

Use github repository to clone all the files.

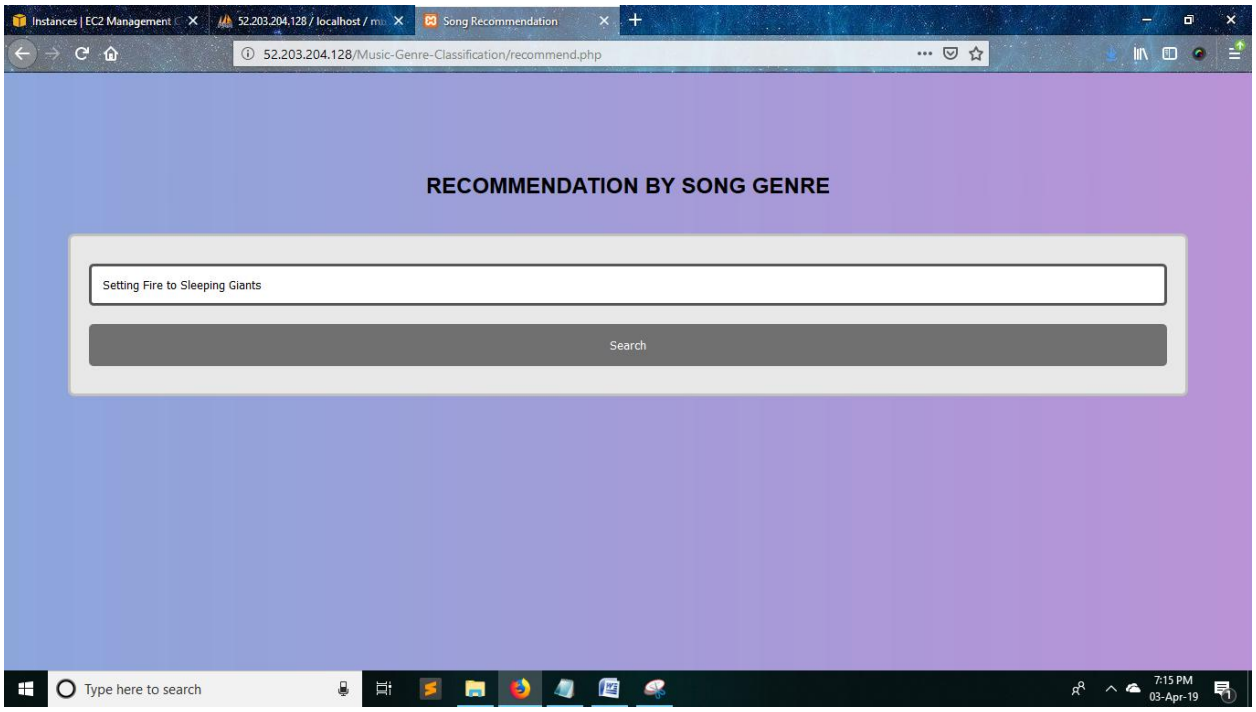
```
sudo git clone https://github.com/Ast97/Music-Genre-Classification.git
```

Use the link http://52.203.204.128/Music-Genre-Classification/ to run the web app.

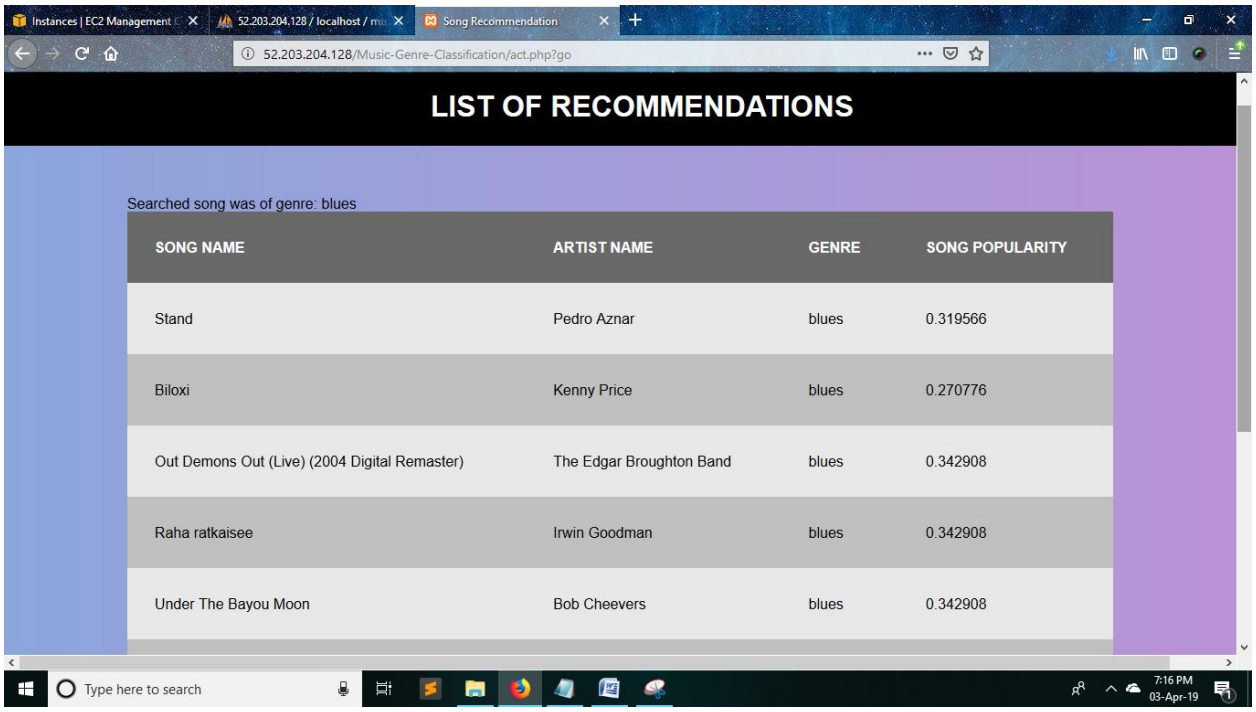



Instances | EC2 Management
52.203.204.128 / localhost / m
Song Recommendation
52.203.204.128/Music-Genre-Classification/
TRENDING SONGS
SONG NAME
ARTIST NAME
GENRE
SONG POPULARITY
Then He Said_ "Sing
Ivan Parker
rap
1.000000
Crayon Sharpener
themselves
blues
1.000000
Daybreak
Winterlight
blues
0.997758
Die Romantic [Live]
Aiden
rock
0.984347
Along Comes Mary
Cal Tjader
pop
0.979837
Give U My Heart
Toni Braxton Featuring Babyface
blues
0.972387
Type here to search
7:15 PM
03-Apr-19
Instances | EC2 Management
52.203.204.128 / localhost / m
Song Recommendation
52.203.204.128/Music-Genre-Classification/
TRENDING ARTISTS
ARTIST NAME
ARTIST POPULARITY
Atom
1.082503
Chanta Vielma
1.082503
Esa Pakarinen
1.082503
Jennifer Lopez
1.082503
Plump DJs
1.082503
Rick Snel
1.082503
Type here to search
7:15 PM
03-Apr-19

Type in a song name.



10 random song recommendations will appear based on the genre of the song entered.



Challenges Faced

- The biggest caveat of the MSD is that you don't have the actual audio to work with. You only have the metadata and audio features as computed by the Echo Nest. The features they provide are pretty good, but they may not be adequate for every application or classification task. If you decide to use the dataset you are limiting yourself to their method of feature extraction and specific machine learning methods. You would not be able to perform an unsupervised feature learning method like deep learning.

- When creating and training an ML model, the goal is to select the model that makes the best predictions, which means selecting the model with the best settings (ML model settings or hyperparameters). In Amazon Machine Learning, there are four hyperparameters that you can set: number of passes, regularization, model size, and shuffle type. However, if you select model parameter settings that produce the "best" predictive performance on the evaluation data, you might overfit your model. Overfitting occurs when a model has memorized patterns that occur in the training and evaluation datasources, but has failed to generalize the patterns in the data.

  This problem can be tackled by splitting the training data into training data and validation data. This avoids the model to have high generalization error and low training error, which signifies model overfitting.

- Managing access to Amazon S3 Resources:
  By default, all Amazon S3 resources—buckets, objects, and related subresources (for example, lifecycle configuration and website configuration)—are private: only the resource owner, an AWS account that created it, can access the resource. The resource owner can optionally grant access permissions to others by writing an access policy.

  To grant permissions for Amazon ML to access one of your S3 buckets, you must edit the bucket policy. To do so we must grant Amazon ML the following permissions to the S3 location where your input data is stored:
  1. GetObject permission on the S3 bucket and prefix.
  2. ListBucket permission on the S3 bucket. Unlike other actions, ListBucket must be granted bucket-wide permissions (rather than on the prefix). However, you can scope the permission to a specific prefix by using a Condition clause.

- Cross platform OS -
  Developers working on the same project often share resources and communicate to achieve the result. The challenge often faced is the diversity of the environment - underlying Operating System - they individually work on. The issues faced were:
  - XAMPP - Configuration of the PHP files and the PHPmyadmin.
  - MySQL - server connectivity, admin and user privileges were different.

- XAMPP - with the recent releases of XAMPP (irrespective of the OS environment), the functioning is hindered by access grant issues to the localhost. The issues can be resolved by tweaking the conf file of Xampp, as stated above (refer to Pg No. 251).
- Troubleshooting Connecting to Your Instance - refer to https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/TroubleshootingInstancesConnecting.html

- SSH configuration with PuTTY – The Host Name/ IP address required in PuTTY configuration can be found in the connect dialog box. The converted private key (.ppk) file is opened for the authentication.