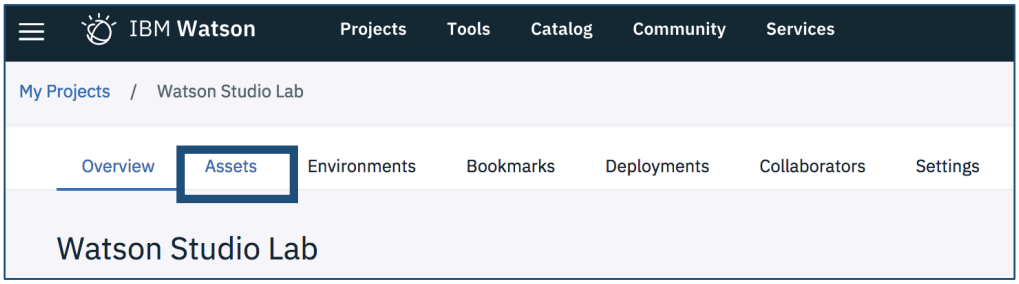
# Lab 1: Model Builder

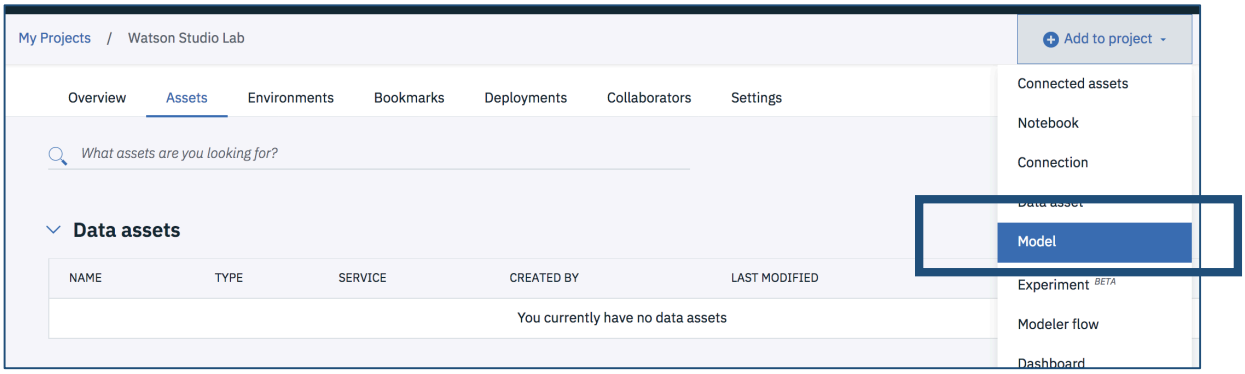
The model builder in IBM Watson Studio guides you, step by step, through building a model that uses popular machine learning algorithms. Just upload your training data, and then let the model builder automatically prepare your data and recommend techniques that suit your data.

## Step 1: Add model to project

- From within your Watson Studio Lab project, move to the **Assets** tab



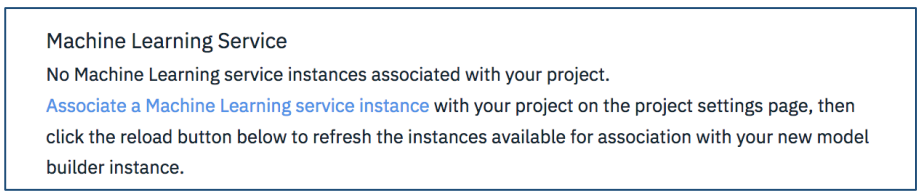

- From the top right, click **Add to project**, and select **Model**



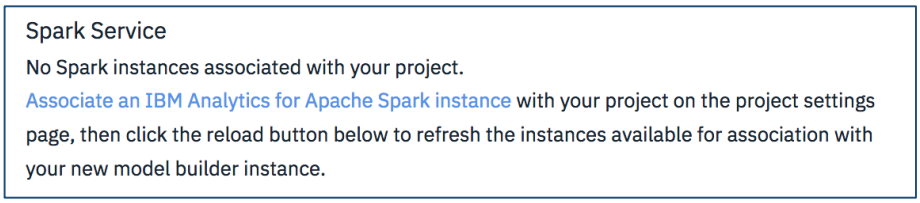

We will be using a dataset which contains telco subscriber data, and movies that the subscriber has streamed in the past. This data can be used to provide movie recommendations for new users.

- Name the model **'Movie Prediction'**

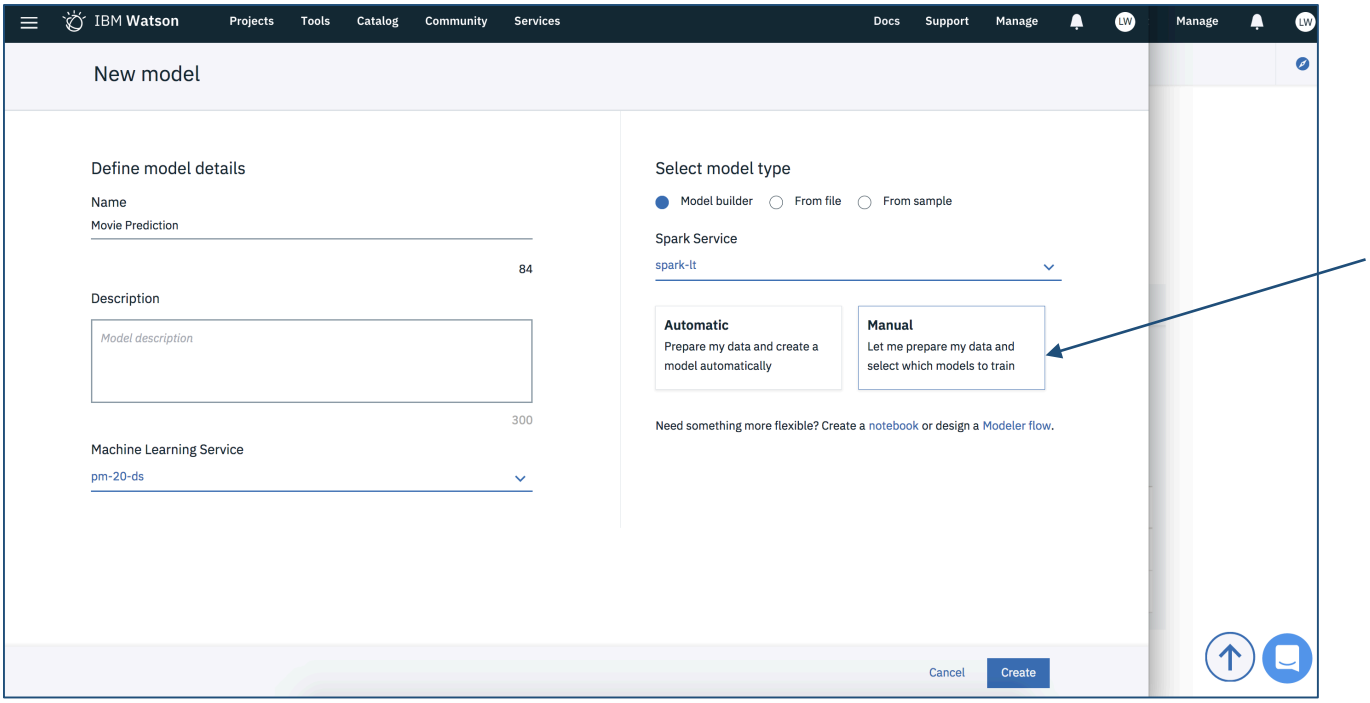
- To proceed, we need to associate a machine learning service with our project (similar to how we created our Cloud Object Storage instance). Click **Associate a Machine Learning service instance**



- Choose the **Lite** plan, then click **Create** and **Confirm**
- Click **Reload** to associate the service you just created with the model
- You'll need to repeat the above steps to do the same for creating an Apache Spark instance




- Once you have your Machine Learning and Spark services associated with the project, select **Manual**, and click **Create**. Your screen should look like the following before you hit Create



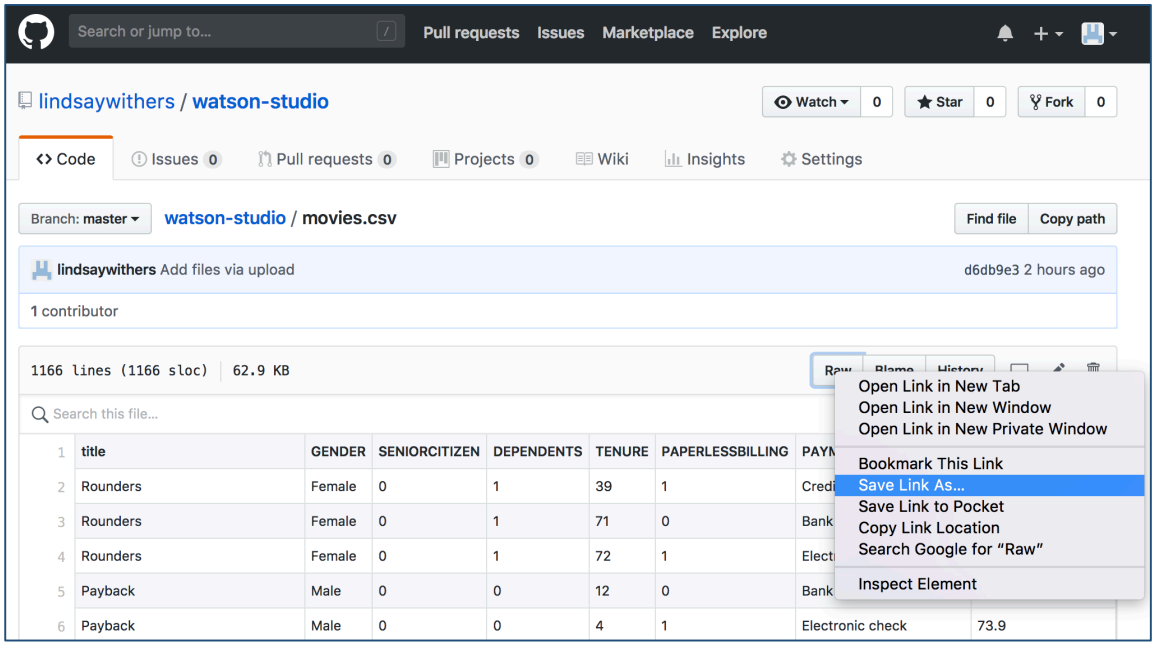

## Step 2: Supply training data

You can supply your structured, historical data in one of two ways:

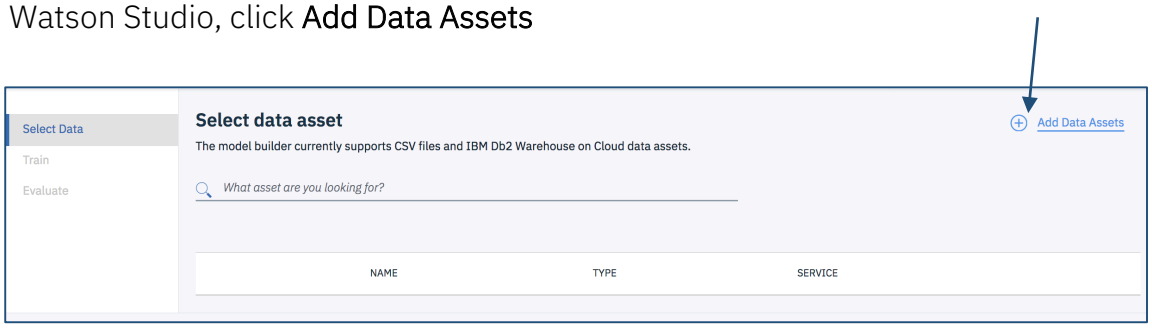
- Upload a .csv file
- Connect to an IBM Db2 on Cloud database

For now, we will upload the **movies.csv** file found in the Git repo: https://github.com/lindsaywithers/watson-studio

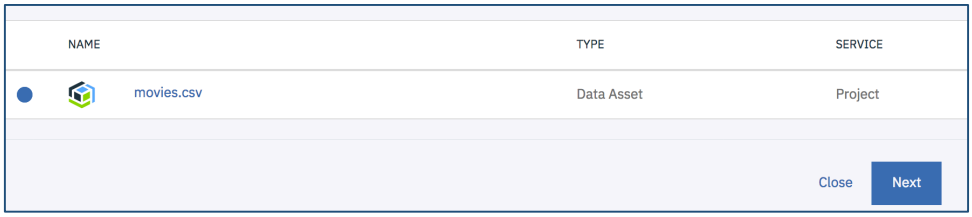
- Open the movies.csv file from the Git repo
- Right-click **Raw** and **Save Link As** to save the CSV to your desktop. **Make sure you save it as a CSV**




- In Watson Studio, click **Add Data Assets**




- From the **Load** tab, click **browse** and select the movies.csv file you saved to your desktop
- Select the data asset and click **Next** to load in the data



## Step 3: Create & Train

- Select **'title'** to be your Label column. Notice that Watson will provide a suggested technique given your data and what it is that you're trying to predict. In this case, our label column (movies) contains multiple distinct values, so multiclass classification is the appropriate technique

- At the top right, you can choose one or more specific estimators

Estimators available when you choose the multiclass classification technique

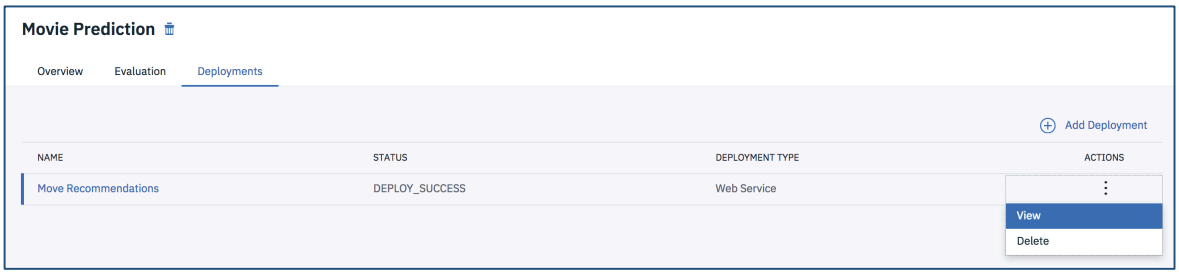
Table 3. Estimators you can assemble together to manually build a multiclass classification model in the model builder

| Estimator | Description |
|---|---|
| Decision tree classifier | Maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It supported both binary and multiclass labels, as well as both continuous and categorical features. |
| Random forest classifier | Constructs multiple decision trees to produce the label that is a mode of each decision tree. It supports both binary and multiclass labels, as well as both continuous and categorical features. |
| Naive Bayes | Classifies features based on Bayes' theorem, which assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. |

- You may select one, or all for training/evaluation

- Once the model(s) are evaluated, and you are satisfied with the performance – select one and click **Save**

- To deploy the model and use in a production application, move to the deployment tab and click **Add Deployment**
    - On the **Deploy model** page, select the **Online** deployment type and type a deployment name
    - Click **Deploy**
    - When model deployment is complete, from the **Actions menu**, click **View**.



- The **Deployment Details** window appears. Note the scoring end point for future reference