# Machine Learning for Dummies with Apache Spark on Bluemix

**Before we begin:**

* Go to datascience.ibm.com and sign up for Data Science Experience. If you have an account already, log in.

**Let’s get started!**

Machine learning is the study of constructing and using algorithms that can learn from and make predictions based on data through building a model from sample inputs.

Some examples of machine learning in action are product recommendations based on users purchase history, pricing for houses based on known prices and variables (or features), and clustering user behavior.

In this lab we will go through 1-3 implementations of machine learning. The first is an introduction, classifying passengers on the titanic based on demographics. The second is a practical application, making product recommendations based on user purchase history. The third is an interesting use case, clustering users’ comments based on comment topic through unsupervised learning.

1. Once you’re logged into Data Science Experience, create a new project. (Click the three bars in the top right, click my projects, and then create project). You can name it whatever you’d like.
2. There aren’t any notebooks or data assets in your project yet. Add the two data files in the box folder (OnlineRetail.csv.gz and train.csv)
3. Now you can start your analysis. After adding the data, click on “add notebooks”, then click on the “From URL” tab. Name the notebook, and paste any of the notebook URL’s in the box folder.
4. From here, just follow the code! Shift+Enter runs the current cell you have highlighted. If you see “XXXXX” anywhere in the code, you need to enter your Object Storage credentials. To see what they are, click the “Find and Add Data” tab on the top right, find the data file you’re looking for, and click “Insert to code”.

* <https://raw.githubusercontent.com/rosswlewis/WorldOfWatsonSparkML/master/Titanic%20Classification%20World%20of%20Watson.ipynb>
* <https://raw.githubusercontent.com/rosswlewis/WorldOfWatsonSparkML/master/Retail%2BCollaborative%2BFiltering%2BWorld%2Bof%2BWatson.ipynb>
* <https://raw.githubusercontent.com/rosswlewis/WorldOfWatsonSparkML/master/Reddit%20TF-IDF%20and%20Clustering%20World%20of%20Watson.ipynb>
* <https://raw.githubusercontent.com/rosswlewis/WorldOfWatsonSparkML/master/Reddit%20Visualization%20World%20of%20Watson.ipynb>