MACHINE LEARNING MODEL DEPLOYMENT WITH IBM WATSON STUDIO CODE

Phase:3 -Development part 1

Project Overview:

Using IBM Watson Machine Learning, you can deploy models, scripts, functions, and web apps, manage your deployments, and prepare your assets to be put into production and to generate predictions and insights.

The Watson Machine Learning service is not available by default. An administrator must install this service on the IBM Cloud Pak for Data platform. To determine whether the service is installed, open the Services catalog and check whether the Watson Machine Learning service is enabled.

Watson Machine Learning provides a full range of tools and services so that you can build, train, and deploy Machine Learning models. Choose the tool with the level of automation or autonomy that matches your needs, from a fully automated process to writing your own code.

Project Objectives:

- Watson Studio helps you inject decision intelligence into your applications with the combined power of predictive and prescriptive analytics.
- This seamless collaboration in a unified environment leads to substantial productivity gains that save both time and money in building, deploying and managing AI models.
- Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data.
- Deployment is one of the last stages in the machine learning life cycle and can be one of the most cumbersome.
- Deploying a machine learning model, known as model deployment, simply means to integrate a machine learning model and integrate it into an existing production environment, where it can take in an input and return an output.

- Watson Machine Learning provides a full range of tools and services so that you can build, train, and deploy Machine Learning models

- Watson Studio allows you to train, deploy, and manage your AI models, and prepare and analyze information during a single integrated environment.

- Watson Knowledge Catalog drives collaboration and transforms information and AI into a trusted enterprise resource through dynamic data policies and requirements.

- IBM Watson Studio helps us to extract the data from documents in a very effective way and it uses key value pairs and extract the data automatically.

- A cloud deployment model defines the cloud services you are consuming and the responsibility model for who manages them.

- It defines your cloud architecture, scalability of your computing resources, what you can change, the services provided to you, and how much of the build you own.

- Watson Machine Learning supports popular frameworks, including: TensorFlow, Scikit-Learn, and PyTorch to build and deploy models.

- Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data.

- The API uses machine learning to automatically extract data from your documents and images and then analyze it to find patterns and relationships. The API can also be used to create custom models to analyze your data in new ways. The Discovery API is a great way to get started with machine learning and is easy to use.

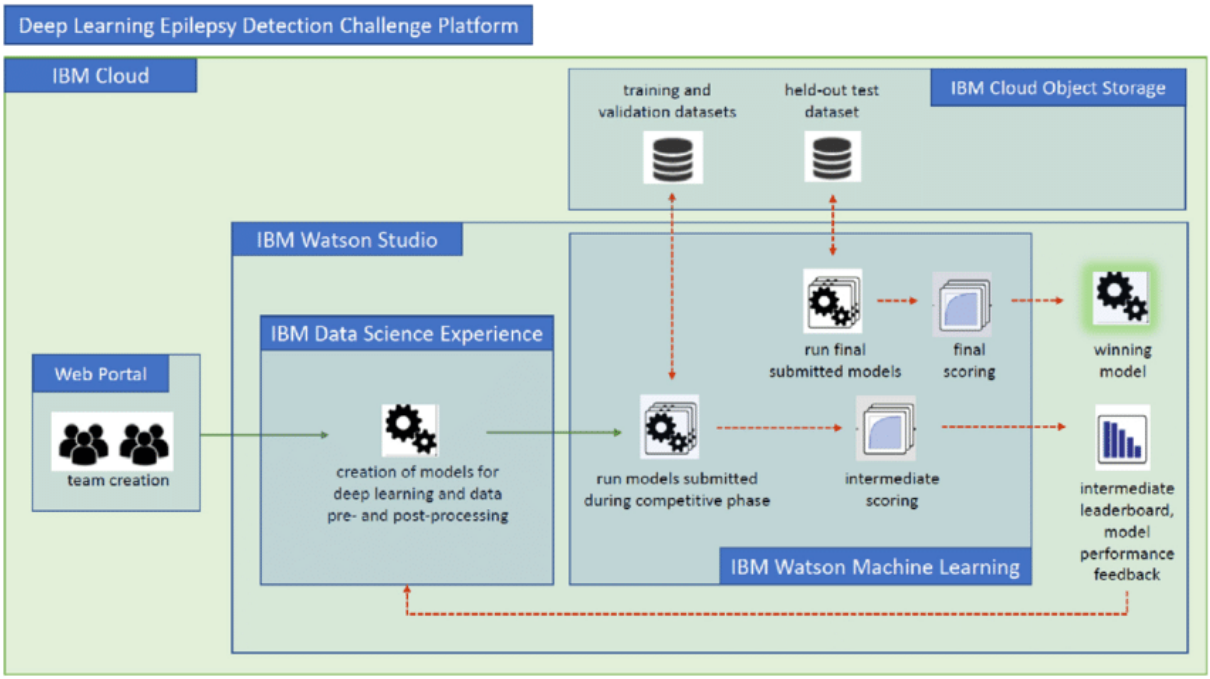
Hardware components:

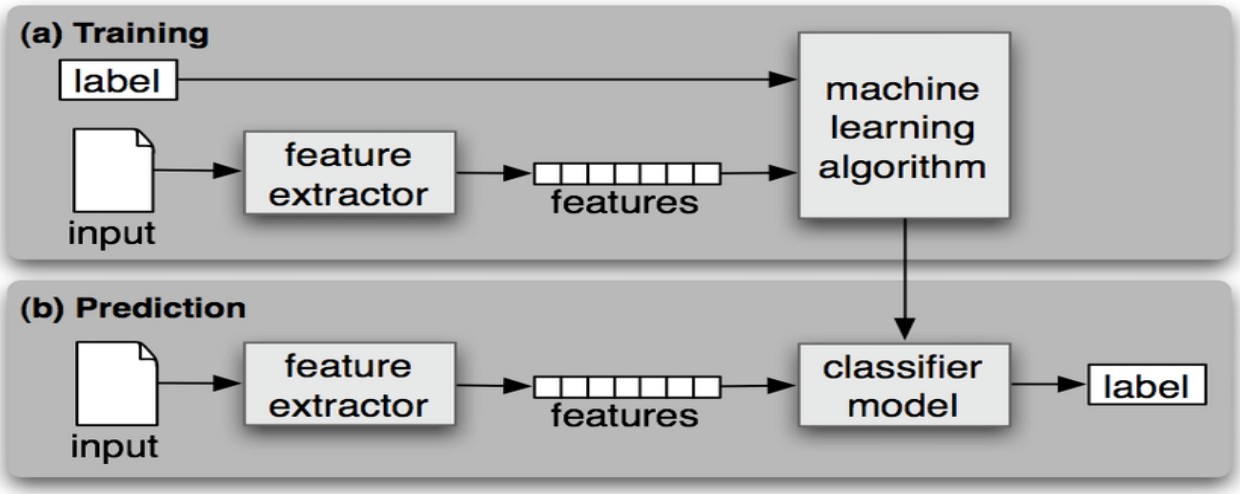
Requirement	Management hosts	Compute hosts	Notes
RAM	64 GB	32 GB	In general, the more memory your hosts have, the better performance is.
Disk space to extract install files from the WML Accelerator install package	16 GB (First management host only)	NA	
Disk space to install IBM Spectrum Conductor™	12 GB	12 GB	
Disk space to install IBM Spectrum Conductor Deep Learning Impact	11 GB	11 GB	
Additional disk space (for Spark instance group packages, logs, and so on.)	Can be 30 GB for a large cluster	1 GB*N slots + sum of service package sizes (including dependencies)	Disk space requirements depend on the number of Spark instance groups and the Spark applications that you run. Long running applications, such as notebooks and streaming applications, can generate huge amounts of data that is stored in Elasticsearch. What your applications log can also increase disk usage. Consider all these factors when estimating disk space requirements for your production cluster.

Software components:

Hardware	Operating system	GPU software
POWER8	Red Hat Enterprise Linux (RHEL) 7.6 (ppc64le)	• CUDA Deep Neural Network (cuDNN) 7.5 library • NVIDIA CUDA 10.1

		• NVIDIA GPU driver 418.40 • NVIDIA NCCL2 2.4 • Anaconda 2018.12
POWER9 with this security fix: RHSA-2018:1374 - Security Advisory	RHEL 7.6 (ppc64le)	• CUDA Deep Neural Network (cuDNN) 7.5 library • NVIDIA CUDA 10.1 • NVIDIA GPU driver 418.40 • NVIDIA NCCL2 2.4 • Anaconda 2018.12
x86	RHEL 7.6	• CUDA Deep Neural Network (cuDNN) 7.5 library • NVIDIA CUDA 10.1 • NVIDIA GPU driver 418.40 • NVIDIA NCCL2 2.4 • Anaconda 2018.12





Tools used to deploy ML model in Watson Studio:

- Tensorflow
- PyTorch
- Scikit-Learn

Tensorflow:

- Tensorflow is an end to end open source platform for machine learning.

- Tensorflow is a rich system for managing all aspects of a machine learning systems.
- The class focuses on using a particular tensorflow API to develop and train machine learning models.
- **Uses:**

  The Tensorflow platform helps you implement best practices for data automation, model tracking, performance monitoring and model retaining.

PyTorch:

- PyTorch is an open source machine learning framework based on the python programming language and the torch library.
- It is primarily developed by Facebook's AI research group.
- PyTorch providing frictionless development and easy scaling through pre-built images, large scale training on GPUs, ability to run models in a production scale environment.
- **Uses:**

Torch is an open source ML library used for creating deep neural networks and is return in the Lua scripting language.

Scikit-learn:

- Scikit-learn is an open source data analysis library, and the gold standard for Machine Learning (ML) in the Python ecosystem.
- Key concepts and features include: Algorithmic decision-making methods, including: Classification: identifying and categorizing data based on patterns.
- Scikit-Learn, also known as sklearn is a python library to implement machine learning models and statistical modelling.
- Uses:

  Through scikit-learn, we can implement various machine learning models for regression, classification, clustering, and statistical tools for analyzing these models.

Conclusion:

All learning involves transfer from previous experiences. Even initial learning involves transfer that is based on previous experiences and prior knowledge. During training, the machine learning algorithm is optimized to find certain patterns or outputs from the dataset, depending on the task. The output of this process - often a computer program with specific rules and data structures - is called a machine learning model.

“PREDICTING FUTURE ISN’T MAGIC, IT’S ARTIFICIAL INTELLIGENCE .”