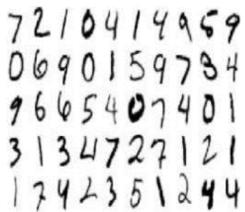
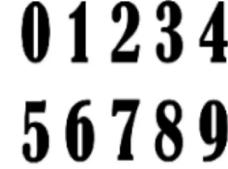
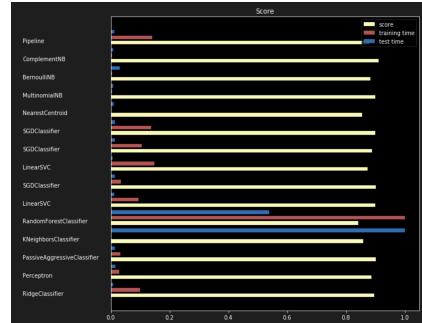


# Banafsheh Hassani

Project	Tools & Methods	Description
<a href="#">Digit Recognition - PyTorch</a>	PyTorch library Apache Spark single node driver CUDA MNIST Dataset Databricks Notebook Convolutional Neural Networks (CNN) Stochastic Gradient Descent (SGD) Docker Google Colab	This project is your key to unlock the power of machine learning. Using the technology behind voice recognition and face detection, it analyzes handwritten numbers to learn and predict. It's like your own digital detective, sifting through data and picking out patterns that would be invisible to the human eye. Running on the cloud, it's fast, efficient, and can handle vast amounts of data. Perfect for businesses that need to understand and use complex information quickly and easily.
<a href="#">Digit Recognition - Tensorflow</a>	TensorFlow, PyTorch torch, torch.nn, torch.optim, torchvision.datasets, torchvision.transform s, torch.utils.data.DataL oader Convolutional Neural Network (CNN), MNIST dataset, data transformation, model training, model evaluation, cross-entropy loss, batch normalization, dropout.	  
<a href="#">Deep Learning: End-to-End House Price Prediction</a>	TensorFlow Keras Hyperopt MLflow Scikit-learn StandardScaler ModelCheckpoint EarlyStopping TensorBoard Flask FastAPI REST API	This notebook demonstrates the process of building an end-to-end deep learning model for house price prediction using TensorFlow Keras, Hyperopt for hyperparameter tuning, and MLflow for experiment tracking and model management. The model is trained on the California Housing dataset from scikit-learn.

## [Text documents classification by sparse features](#)



Python : Scikit-learn  
NumPy Matplotlib  
Jupyter Notebook  
Methods: TF-IDF  
Vectorization Ridge  
Classifier Perceptron  
Passive-Aggressive  
Classifier k-Nearest  
Neighbors Random  
Forest Classifier  
Metrics (Accuracy,  
Dimensionality,  
Density, Classification  
report, Confusion  
Matrix)

This project uses various classifiers to perform text document classification on the 20 newsgroups dataset. It uses TF-IDF for feature extraction, and calculates various performance metrics for each classifier, including accuracy, dimensionality, density, top 10 keywords per class for some classifiers, classification report, and confusion matrix.

## [Text Classification Using SGDClassifier and GridSearchCV](#)

Python  
Scikit-learn Library Logging ,  
NumPy Library Matplotlib  
Library Methods/Techniques:  
Ridge Classifier Perceptron  
Passive-Aggressive Classifier  
K-Nearest Neighbors (KNN)  
Random Forest TF-IDF  
Vectorization Training-Testing  
Split Accuracy Scoring  
Classification Report  
Confusion Matrix Feature  
Selection (Top 10 keywords  
per class)

This project involves loading a subset of the 20 Newsgroups dataset, performing text feature extraction and using a SGDClassifier to classify the documents. The project also uses GridSearchCV to optimize the parameters of the text feature extraction and the classifier.

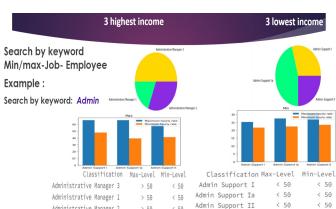


## [MLflow Experiment for Diabetes Prediction with Random Forest](#)

MLflow (for experiment tracking and logging)  
scikit-learn (for dataset loading, model creation, and evaluation)  
Random Forest (as the machine learning algorithm for regression)

This project demonstrates the use of MLflow to track machine learning training runs in a notebook environment. It includes creating and training a Random Forest model on a diabetes dataset, logging model parameters and metrics, and organizing runs within an experiment.

## Job Analysis



Data processing  
Problem statement  
Data cleaning  
Exploratory analysis  
Feature engineering  
Methodology

**Output:** Get user expectation (by asking questions from user) and recommend some sorted top 3 - tail 3 jobs in that category with salary range/ gender.

## lululemon (SSC) Zoom Room & Printer Tracker



Excel  
Macro  
VBA  
Pivot+  
visualization  
Data entry  
user form

Tracking all zoom rooms and printer on the site live with a live board report

My job at lululemon involves checking and troubleshooting all Zoom rooms and printers on-site. Although they have a tracker, it is not efficient for me. Having a cohesive database allows managers to stay informed about the site's status and find the best solutions for any issues that arise. All actions performed on the system are collected in the database for future analysis. Additionally, this system includes live pivot/visualization boards.

## Equipment tracker + desk status live board

### Desk status Live board.



SQL  
Server-side  
database  
Dynamic web page  
Tableau, Excel, SQL,  
Python, data analysis  
methods  
HTML,CSS,PHP,Javascript

Client-side  
web base  
smart form  
Data entry  
web user  
form  
Live Tableau  
dashboards

This project starts to improve the booking desk system + tracking company workstation equipment (Monitor, keyboard,.). Desk status live report + tracking staff  
Give live insight of site

- ✓ Desk status live report
  - ✓ Tracking staff and equipment
    - -> fill the form -> Check in/out from book setting easier.
    - Advantages
      - Staff Clock in / Clock out
      - Live desk status report
      - Live equipment tracker

Banafsheh8Hassani@gmail.com

+1 236-889-9014

**Vancouver, BC**

## Food order

✓ Business Analysis

✓ Microsoft Visio

✓ ECS System

Use-Cases

- Use-Case Name
- Actor
- Trigger
- Responses

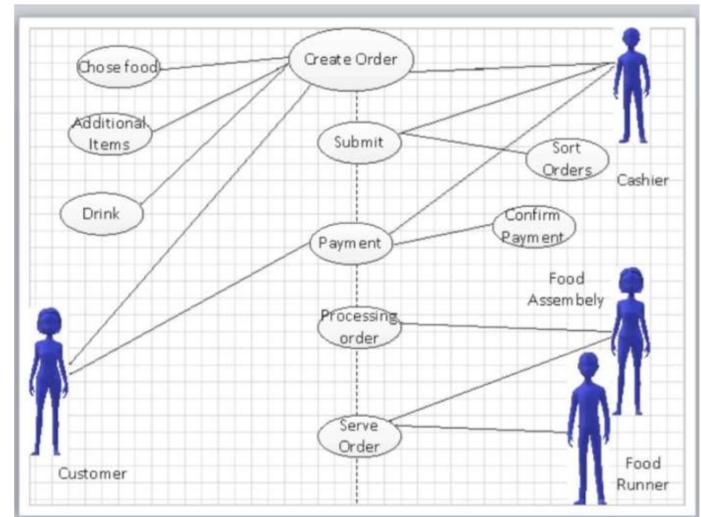
✓ Check-out

Equipment

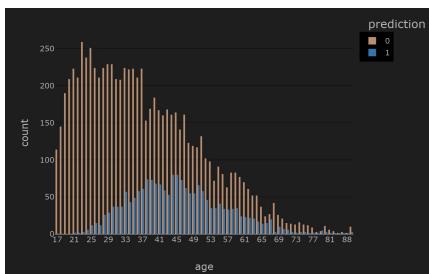
- Actor Action
- System Response.

Food ordering process.

Analysis of food ordering process.



## Predict if an individual's income > \$50,000



MLlib

Hyperparameter tuning

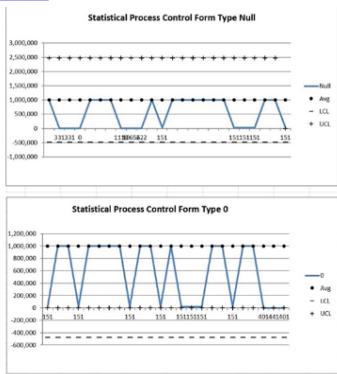
Income > 50K /each age range

✓ Using Apache Spark's ML library

✓ Investigates binary classification problem.

✓ Make predictions and evaluate.

## Greater Vancouver Business License



● Dataset anomalies:

- Blank cells
- Duplicates
- Abbreviations

● Data quality assessment

(DQA)

- Initial assessment:
- SME review, research, suggestions by DQ rules.

[Vancouver Business License dataset](#) includes columns that are relevant to the analysis. Upon initial screening of the dataset, certain anomalies were identified and extensively analyzed. The graph illustrates the Special Purpose Corporation (SPC) data.