Must have skills in pocket for Machine Learning or Data Science

---

1. **GIT CLI Interface**
    - complete workflow from creating github repository
    - automate testing for every push of new code.

2. **Complete Python**
    - Data Types (native and container data types)
    - Object Oriented Programming System (OOPS)
    - Various Inbuilt modules (os, sys, functools, etc.,)
    - Walkthrough of PIP and virtual environment setup

3. **Web Framework of choice (API's Could server the purpose)**
    - Flask (micro service focused)
        - RESTful services
        - JWT token Authentication
    - Django (Little heavy framework but highly recommended)
        - RESTful services
        - JWT token Authentication

4. **Deep Understanding of what is Machine Learning and DataScience**
    - Exploring fields like
        - Structured Data (SQLAlchemy - Python Package)
        - Unstructured Data (PySpark / Spark)
    - Types of Machine Learning Systems
        - Supervised / Unsupervised Learning
            - Regression (supervised)
                - Linear models
                - SVM Regression
            - Classification (supervised)
                - Decision trees
                - Random forests
            - Clustering (unsupervised)
                - K-means clustering
                - KNN
        - Batch and Online Learning
        - Instance and Model based Learning
    - Development cycle of machine learning
        - Collection of data (statistician frame problem first stage)
        - Framing the problem (statisticians collect data in second stage)
        - Select Performance Measure
        - Data Wrangling

        - Exploratory Data Analysis (***Most Imp)
            - Visualization of data
            - Handling missing values
            - Outlier detection and handling them
            - Feature Engineering
        - Building data pipelines
        - Training and evaluating a Model
        - Fine tuning the model hyperparameter (hyperparameter tuning)
        - Deploying and monitoring model
    - Introduction to Deep Learning
        - What is deep learning
        - ANN, CNN and RNN Architectures
        - Tensorflow introduction (python package)

LIST OF PYTHON PACKAGES (We will cover)

- **In Built Libraries**
    - os
    - sys
    - cocnurrent_futures
    - typing
    - json
    - csv
    - logging
    - dataclasses
    - re
    - argparse
    - ast ->abstract syntax trees
    - joblib
    - functools
    - itertools
    - threading --> talk about Queue
    - multiprocessing --> talk about Queue
    - contextlib --> @contextmanager [implement with try: and finally:]
    - pickle
    - configparser
    - unittest
- **Web Scraping**
    - selenium
    - requests
    - bs4 --> BeautifulSoup
    - scrapy
- **Web FrameWorks**
    - Django
    - Bottle
    - Flask
- **DataBase Tools**
    - SqlAlchemy

- **Machine Learning and Data Science**
    - numpy
    - pandas
    - sklearn
    - statsmodels
    - matplotlib
    - seaborn
    - imblearn - for balancing imbalanced datasets (up/down sampling)
    - feature_engine - for feature engineering
    - bokeh
    - boosting
    - bagging
    - xgboost
    - lightgbm (light Gradient Boosting Machine)
    - cat Boost
    - eli5 (explain like i am 5 years) - for debugging machine learning models.
- **Deep Learning**
    - Tensorflow (or) Pytorch
    - keras
- **Image processing**
    - opencv
    - Pillow / PIL
    - pytesseract
    - tesserocr
    - gluonCV
- **NLP**
    - gensim
    - NLTK
    - spacy