Below are some samples of my work product in software engineering, machine learning, data engineering, and data science areas. They are a mixture of proprietary projects, open-source projects, and academic contributions (All the samples are based solely on individual efforts)

1. **Language Identification Model**: I trained and built a supervised machine learning model to identify the name of a language based on the written text, with 98% accuracy. **Techniques**: neural networks/distributed representations of words. **Tools**: fastText, Python, pandas, NumPy, bash scripts
https://github.com/iamaziz/language-detection-fasttext

2. **Datasets Module**: I built a Python package that provides instant access and download for over 750 scientific datasets which largely used in data science and machine learning research (The project has more than 800 stars on GitHub and tens of thousands of downloads).
https://github.com/iamaziz/pydataset

3. **Sentiment Analysis Model:** I trained and built an unsupervised language model based out of a large corpus of Arabic text. **Techniques**: word embeddings, word2vec. **Tools**: genism, word2vec, NLTK, pandas.
https://github.com/iamaziz/ar-embeddings

4. **TVEyes NLP Project:** I designed and built an intelligent tagging system that recognizes the name of financial institutions mentioned in the stream of TV captions. The system tags each institution with valuable information such as its stock market value at the time of mention. **Techniques**: neural networks, named entity recognition NER. **Tools**: Python, genism, SpaCy

5. **Hubble Social Sync (NBCUniversal) system**: I built an end-to-end Extract-Transform-Load ETL pipeline with daily automation and data modeling. The system extracts Ads metrics and build datasets from large social media websites such as Facebook, Twitter, and GoogleAds. **Tools**: Python, pandas, several APIs and their respective packages.

6. **Bayesian Decision Theory**: I implemented a 2D Bayesian classifier based on Bayesian decision theory for non-parametric estimation.
https://gist.github.com/iamaziz/492383e33490f37db177cde108884c25

7. **Information Gain and Entropy of Probability distribution**: I implemented an algorithm for calculate information gain and probability distribution
https://gist.github.com/iamaziz/02491e36490eb05a30f8

8. **Linear Regression model with gradient descent algorithm:** I implemented (from scratch) a linear regression model based on gradient descent optimization algorithm
https://gist.github.com/iamaziz/ea5863beaee090937fd6828e88653f5e

9. **K-means algorithms implementation:** I implemented (from scratch) the clustering algorithm "K_means" using recursive approach.
https://gist.github.com/iamaziz/0786e3de174c79839e42a5926f25acb2

10. **Rendering relational data in graphs (PyGraph):** I built a software package that helps users to easily render and visualize relational data in graphs to facilitate analysis.
https://github.com/iamaziz/pygraph

11. **Read/Write Utility for AWS Cloud Storage Service**: I designed and implemented simplified interface for reading and writing data between local machines and AWS S3 buckets.

12. **Web Scraping and 3D Visualization:** I developed a project for scraping a public website to extract statistical data and 3D visualize the data.
https://nbviewer.jupyter.org/github/iamaziz/soccer-leagues-interactive/blob/master/soccer-interactive-visualization.ipynb

13. **iversions:** I built a tool for helping Python developers to quickly access meta information about the imported software in their programs.
https://github.com/iamaziz/iversions

14. **TermFeed**: I built a Python package for fetching and reading RSS feeds via terminal.
https://github.com/iamaziz/TermFeed

15. **Sorting algorithms comparison**: I implemented (comparison-based) sorting algorithms from scratch with comparing and visualizing their performance.
https://gist.github.com/iamaziz/8b6a961edbf7f9917dde57de301ae8e9

16. **Demo ETL pipeline**: I created a self-contained Extract-Transform-Load data pipeline as a demo with data modeling (pivot methods) and analysis. Based on: pandas and SQLite.
https://github.com/iamaziz/etl

17. **Predicting Air Traffic Delay:** I applied machine learning-based analysis to understand "*Why do flights get delayed?*". Tools used: Keras, CatBoost, XGBoost, pandas, matplotlib.
https://github.com/iamaziz/air_traffic_delay