Algorithmic and Visualization Capabilities for Machine Learning

By: Caelia Thomas, De'Johnna Wright and Joshua Whitaker

Advisor: Imtiaz Ahmed

Sponsor: Capital One

Background

- American bank which specializes in credit cards, auto loans, banking and saving accounts
- Formerly known as Oakstone Financial, later renamed in 1994
- Headquartered in McLean, Virginia
- Interested in algorithmic and visualization capabilities to support machine learning-driven analysis of large and/or complex data sets

Needs

- Efficient analysis Program

- Informative visual representation

- To predict data trends

Benefits

- Increased efficiency in processing data sets
- Increased analysis depth with an effective visualization of the data
- Reduced error rate by using machine learning algorithms.

Problem Statement

The need for the Capital One team is to increase efficiency in processing data sets, effectively visualize data for in depth analysis and reduce the error rate by using machine learning algorithms.

Design requirements

- Must be able to understand and apply various algorithms.

- Must be able to find complex relationships in large data sets

    - Built-in or library supported regression analysis
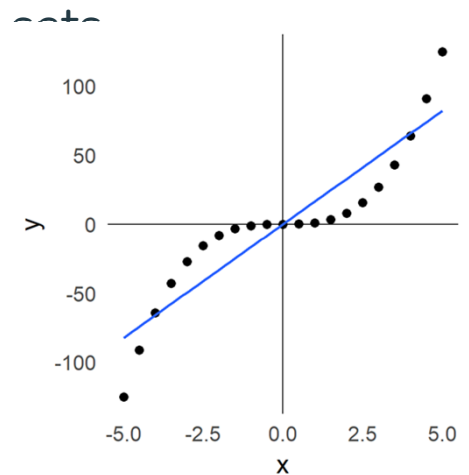
Design requirements (cont.)

- Must produce visuals in a timely manner

    - Execution time of the program should be quick

- Must be able to accurately predict insights based on data sets.

Standards and Regulations

- Compliance with Capital One's NDA agreement

- Workload must fall within 10-hour work week schedule

- Must work in accordance with Dr. Ahmed

- Meet with Capital One once a month (minimum)

Constraints

- Must use Machine Learning approach

    - Continuously learns through experience and use of data

- Must use mostly open source software

    - Avoids copyright issues

Evaluation Criteria

- Data set size

- Data set complexity

- Time used for implementation

- Efficient data visualization

- Selected solution designs: Tensorflow and PyTorch

TensorFlow

Pros:

- Is an open source platform
- Overall better visualization
- More widely used than PyTorch
- Polished interface

Cons:

- More static interface (difficult to use)

PyTorch

Pros:

- Is an open-source platform
- Faster computation times
- More efficient memory management
- Best compatibility with Numpy
- More dynamic interface/control

Cons:

- Newer, less resources available on how to use it

Decision Matrix

|	Visualization Capabilities (0.5)	Efficiency (0.3)	Learning Difficulty (0.1)	Data Handling Capability (0.2)	Total
Pytorch	3	4	4	3	**3.7**
Tensorflow	5	3	2	3	**4.2**

Top Solution Design:
TensorFlow

Detailed Description

- The data set client information is provided from Capital One

- Data set will be cleaned, analyzed and implemented into ML model using TensorFlow

  - A cleaned data set has its missing or improperly formatted fields corrected

- A visual representation will be generated and used for approximating loan eligibility

Block Diagram
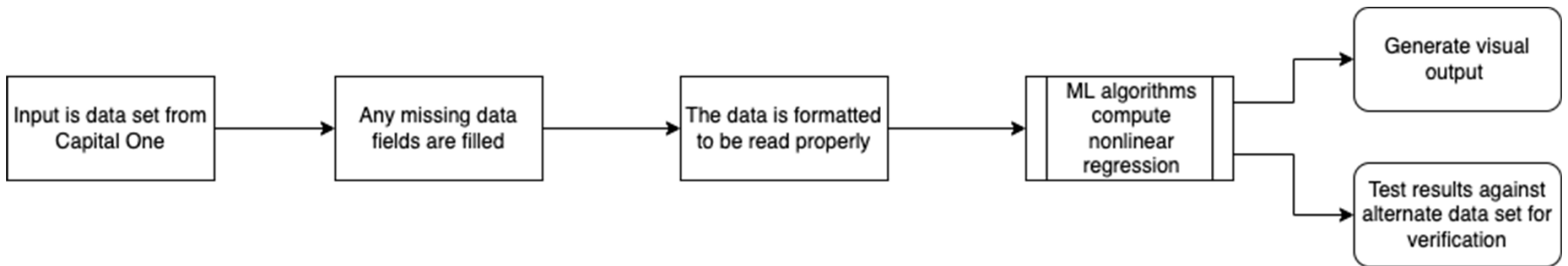
Final Product

- TensorFlow program can handle varied customer data sets

- Visual display provides an accessible way of identifying trends in data
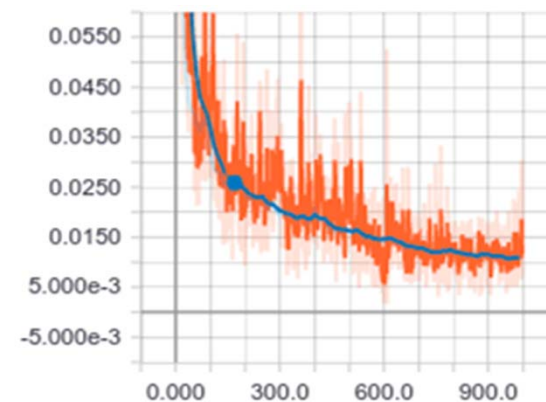
Hypothetical Application

- Data set contains name, salary, savings, financial history... etc

- Regression algorithm evaluates if customers are eligible for a loan

- An efficient visual can provide an accurate prediction as well

Future Works

- Specific technical details will be given in December

- Next Semester (if similar to hypothetical)

    - Cleaning data sets

    - Testing the practical use of TensorFlow algorithm

    - Finding the specific nonlinear regression

Conclusion

- Capital One handles large amounts of customer data
- Providing visualization of the data increases efficiency of analysis
- TensorFlow is a commonly used open source machine learning library that computes nonlinear regression
- Although somewhat harder to learn, TensorFlow's visualization capabilities outperform its alternatives

Thank you for your attention!
Any questions?