# IMPLEMENTING CUSTOMER CLASSIFICATION WITH PYTHON.

Alladin Al-Gannam F18030115
Anna Akulova F19030132

# Why this topic?

The goal of this project is to first, validate that a NN model is more powerful in accuracy than other models and two, how we can leverage this information to mitigate customers from leaving and
reclaim customers that have left SIS.

**IDEA**

- With the rise of information technology, more and more people are adopting the use of e-commerce, especially recently.
- All this leads to a huge number of clients who need to be classified to optimize  business and trade. To solve this problem, we can use the latest advances in artificial intelligence.

**REFLECTIONS**

- Originally, we were thinking of using a readily available dataset and running a deep learning model. While this would have been a decent start, We did not think it would be a real case as compared to what we would encounter as a consultant.

**RESULT**

- Having studied the market and using personal acquaintances, we decided consider improving customer experience and build a neural network (NN) to see if it will a viable model compared to other more traditional algorithms and methods.

# The Jupyter Notebook combines two components:

We decided to do our program in 2 parts: the web server to interpret the graphic result and the python code of our neural network to work with the dataset. As the web server backend, we will use Jupyter Notebook engine, hich will allow us to create GUI environment for our python code of the program. The Jupyter Notebook extends the console-based approach to interactive computing in a qualitatively new direction, providing a web-based application suitable for capturing the whole computation process: developing, documenting, and executing code, as well as communicating the results.

## A WEB APPLICATION

There are many variations of passages of Lorem Ipsum available, but the majority have suffered alteration in some form, by injected humour, or randomised words which don't look even slightly believable.

## NOTEBOOK DOCUMENTS

There are many variations of passages of Lorem Ipsum available, but the majority have suffered alteration in some form, by injected humour, or randomised words which don't look even slightly believable.

# Main UI features of our web application

## AMPLE OPPRTUNITIES

In-browser editing for rich text using the Markdown markup language, which can provide
commentary for the code, is not limited to plain text.

## USABILITY

In-browser editing for code, with automatic syntax highlighting, indentation, and tab completion/introspection.

## FUNCTIONALITY

The ability to execute code from the browser, with the results of computations attached to the code which generated them.

## UTILITY

The ability to easily include mathematical notation within markdown cells using LaTeX, and
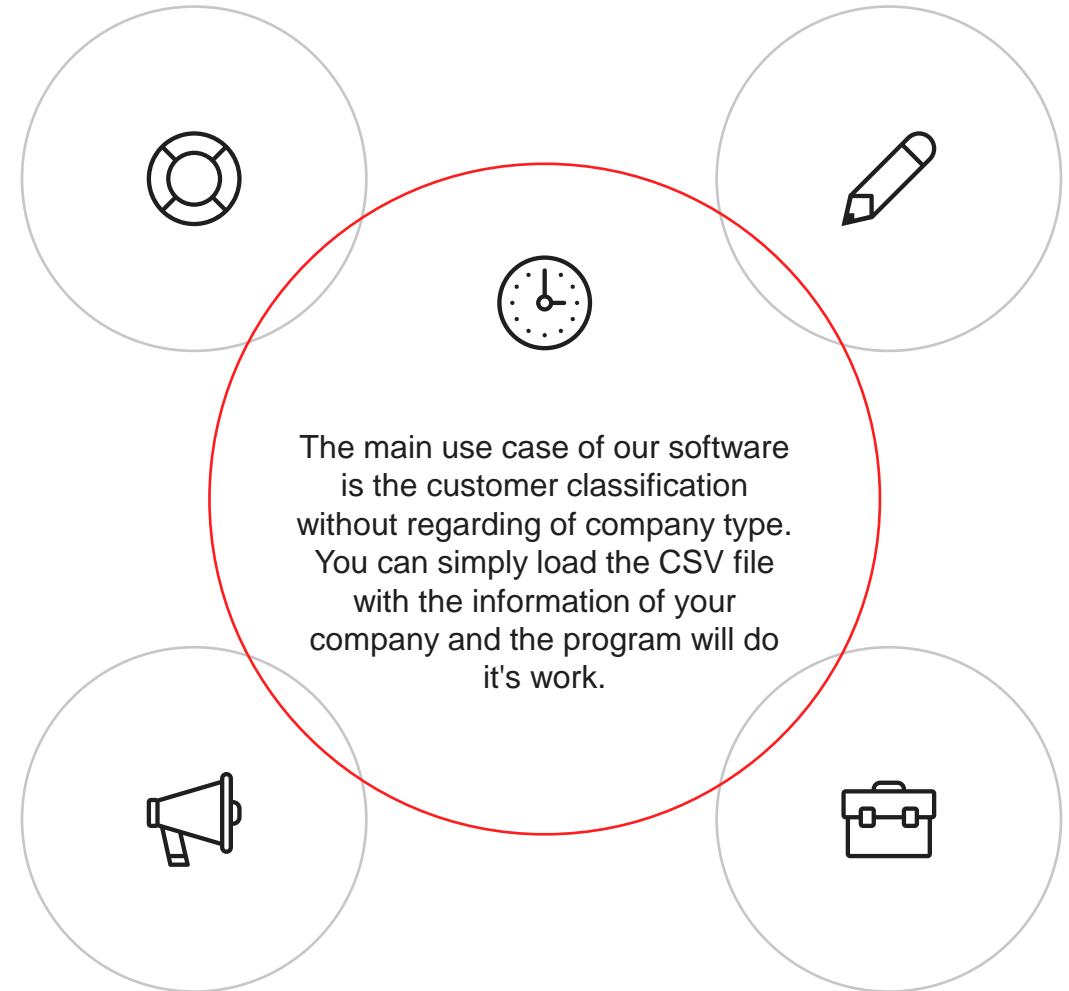rendered natively by MathJax.

## UNIVERSALITY

Displaying the result of computation using rich media representations, such as HTML, LaTeX, PNG, SVG, etc. For example, publication-quality figures rendered by the matplotlib library, can be included inline.

# Why our program is important?

The main use case of our software is the customer classification without regarding of company type. You can simply load the CSV file with the information of your company and the program will do it's work.

Provided Features.

| Line of Business/ Non-Premium | Is Multi-Entity | Policy Status | Active Customer | Customer Number | Type of Business | New Business/ Renewal Group | Total Cost |
|---|---|---|---|---|---|---|---|
| Policy Effective Data | Policy Expiration Data | Customer Address 1 | Customer Address 2 | Customer City | Customer State | Customer Zip Code | Cancel Date |
| Cancel Reason | Writing Company | Policy Number | Invoice Month | Invoice Year | First Written Date | Invoice GL Month | |

The main use case of our software is the customer classification without regarding of company type. You can simply load the CSV file with the information of your company and the program will do/ it's work.

We created new features from the dataset that was provided and formatted the data, so each observation is associated with that customer.

Selected And Newly Created Features.

| Customer ID | Gender | Referrer | State | Paid Full Premium Before | Became Cust | Duration As Cust | Accident/ Health (P) Duration |
|---|---|---|---|---|---|---|---|
| Accident/ Health (P) Amount | Builders Risk (P) Duration | Builders Risk (P) Amount | Dwelling Fire Duration | Dwelling Fire Amount | Earthquake (P) Duration | Earthquake (P) Amount | Flood Duration |
| Flood Amount | Homeowners Duration | Homeowners Amount | Life (P) Duration | Life (P) Amount | Motorcycle Duration | Motorcycle Amount | Private Passenger Auto Duration |
| Private Passenger Auto Amount | Umbrella (P) Duration | Umbrella (P) Amount | Total Duration | Total Amount | Still Customer | | |
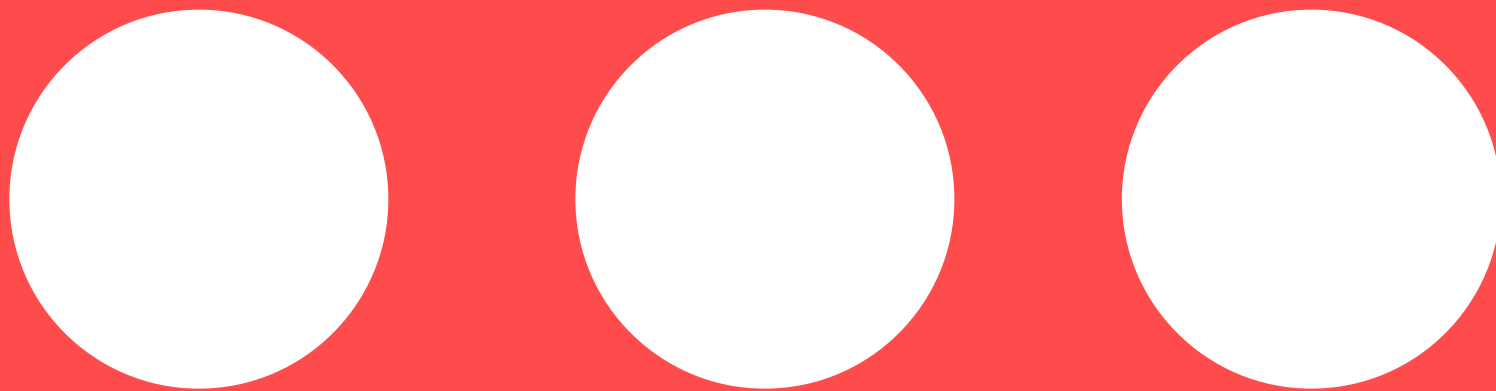
(The "DEMO" visual part: showing the work of the program directly)

# The optimized model and employed an early stop mechanism

Unseen holdout dataset consists of 36 (5%) of the original 713 observations. We executed the optimized model and employed an early stop mechanism.

The Bayesian optimized deep learning model is much better than the previous models and should be used on the client's dataset. Another metric that should look at is the F1 score for selected model. F1 score is the harmonic mean of Precision and Recall. It provides a better measure of the incorrectly classified cases than the accuracy metric. The F1 score for the optimized model 84.6% (Precision: 75.9% and Recall: 95.7%). The F1 score is much higher than Logistic Regression model (79.7%).



Loss Results by Epoch for Unseen Holdout Dataset.
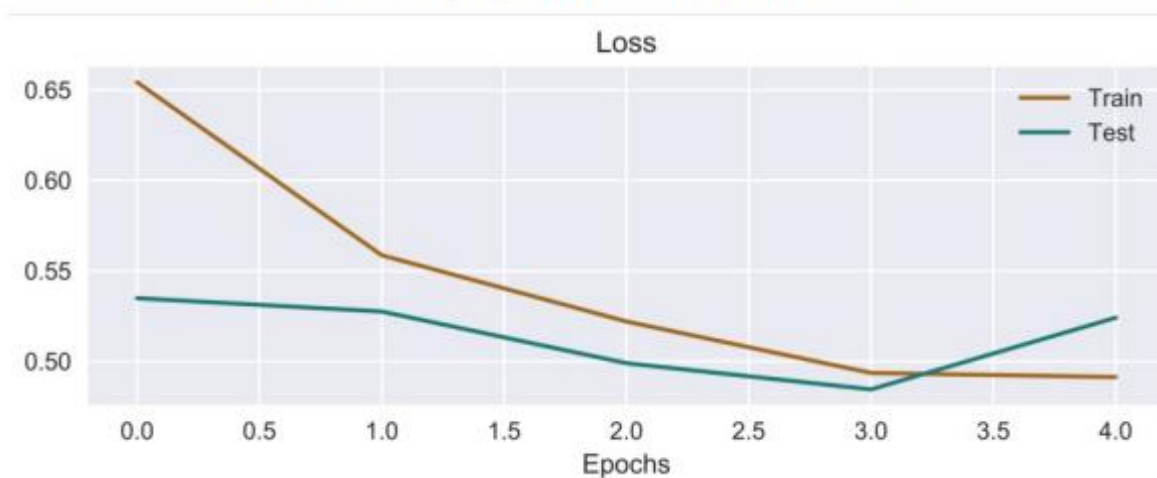


Accuracy Results by Epoch for Unseen Holdout Dataset.
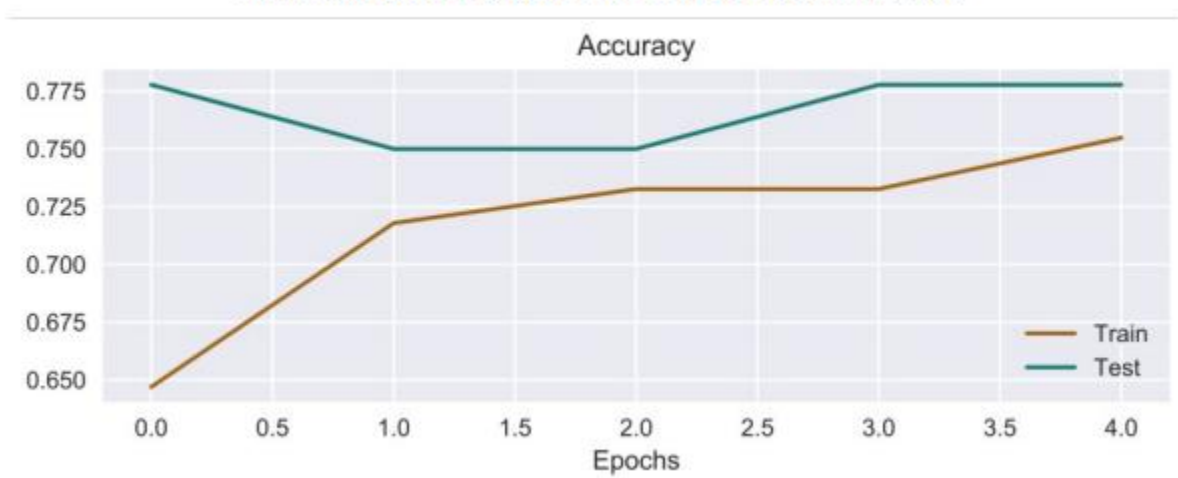
The accuracy results from this optimized model using the unseen holdout dataset
Below is the summary of our accuracy results from the 3 models:
• Supervised Learning (Logistic Regression): 73.6%
• Baseline Single-Layered NN: 75%
• Bayesian Optimized Deep Learning NN: 77.8%

**73.6%**
Supervised Learning

**75%**
Baseline Single-Layered NN

**77.8%**
Bayesian Optimized Deep Learning NN
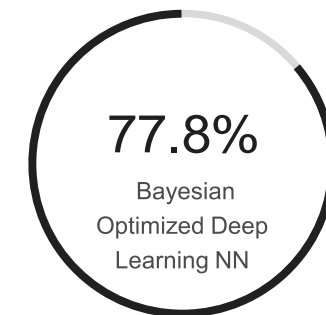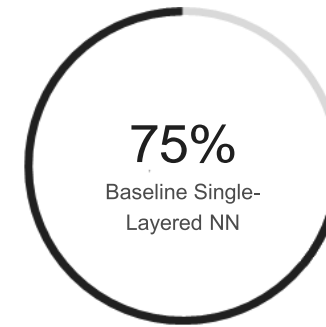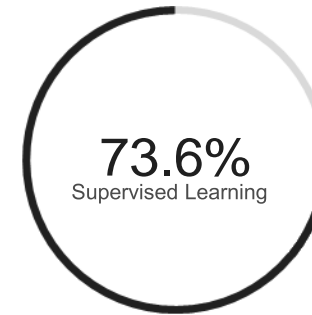
# In the matrix is the low precision
(high number of False Positives)

The figure below shows the confusion

matrix of results.

Confusion Matrix of Bayesian Optimized Deep Learning Model.
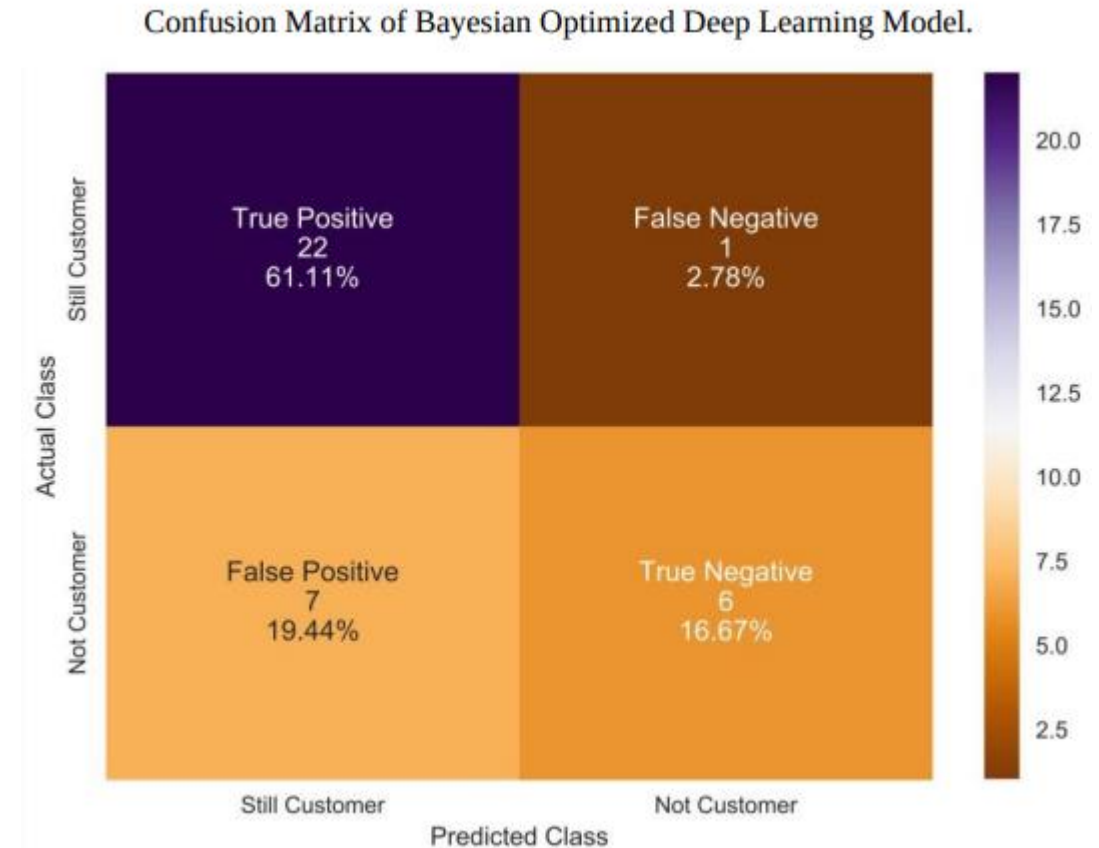
# OUR SERVICE &CONCLUSION

During this project, our team members were able to improve their **programming skills**, in particular programming in python, deep learning and neural networks, and, of course, **software engineering**. Moreover, changing the topic of our work also broadened our horizons in marketing and **management** and gave a new breath to our group.

As a result, we managed to create a **potentially commercially successful** project that is necessary and has no close analogues today. We were able to disassemble the working concept, went deeper into **classification and analysis**, using the method of referring to the **latest digital technologies** in the field of **artificial intelligence**
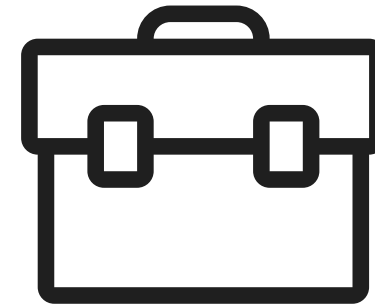
## BUSINESS

Retaining existing customers is big but bringing the customer back to the company is even bigger.
Optimizing and executing models it is helps the company