ACTIVATE
VIRTUAL SEARCH & AI CONFERENCE
Powered by Lucidworks

Life Insurance Risk Prediction

Jerry Thomas, Senior Data Scientist, thinktum

Issue Statement

Issue Statement

- Insurers are analyzed for risk through an underwriting process supported by specialists.
- For an application to be considered, an insurer must submit multiple documents including medical records.
- Then, the underwriting process begins, and risks are determined. This involves various specialists and can be quite expensive.
- Deep learning could help reduce the time it takes to calculate and fasten things up.

Solution

Solution

Supervised Learning

- We use labeled datasets here.
- These datasets are designed to train and predict the outcome accurately.

We ask our underwriters or our group of specialists to give each applicant a risk score.

Unsupervised Learning

- We use unlabeled datasets here.
- These datasets are designed to discover hidden patterns, without any human intervention.

We ask our underwriters or our group of specialists to look at the hidden groups determined by the algorithm and associate with a risk score.

Dataset

Dataset

- For the purpose of demonstration, I have used *Prudential-life-insurance-assessment dataset* obtained from https://www.kaggle.com/c/prudential-life-insurance-assessment/data.
- It is a supervised dataset labeled by experts for life insurance risks.
- This dataset has 50k + observations and 128 features for each observation.
- Each organization can create its own features that include numerical, categorical, textual, and image data. For the purpose of this demonstration, I have limited the dataset to numerical and categorical data.

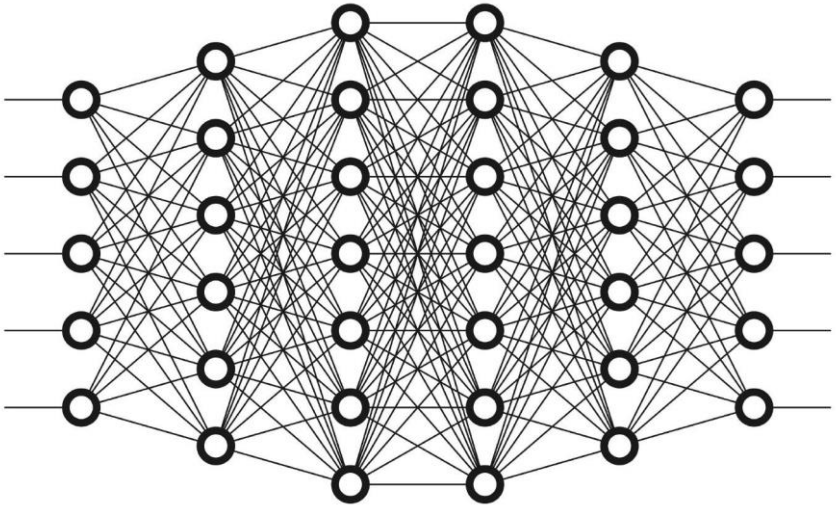
What is Deep Learning?

What is Deep Learning

- An artificial neural network model is used to determine the risk of the applicant. Designed after the human brain, this system mimics the functions of the brain.
- The network we are using is a multilayer perceptron.
- Algorithms in the hierarchy apply nonlinear transformations to their inputs and create statistical models from their learnings. Until the output is accurate enough, iterations are conducted.

The Model

The Model



- The next step is to build a deep learning model using the extensive information from the dataset.
- The demonstration would use a simple deep learning model with 5 layers and 1024, 512, 256, 128 and 64 nodes respectively for each layer.

Results

Results

- Performance is evaluated after training the model to check if our risk model is helpful to evaluate the applications. For comparison purposes, I have also trained an XGBoost model on the data. The table below summarizes the results.

Deep Learning	Mean Absolute Error	Deviation
Absolute Error	1.6	1.3
Relative Error	0.2	0.18
Percentage Error	22.7	16.5

XGBoost	Mean Absolute Error	Deviation
Absolute Error	2.6	1.5
Relative Error	0.5	0.22
Percentage Error	35.7	18.5

Analysis

Analysis

- Deep learning is not a black box anymore.
- Techniques like "Lime" (Local Interpretable Model-agnostic Explanations) can be employed to explain the model's behavior and prediction.

Using these techniques, we can see that **BMI**, **weight** (Wt), and **medical history** are the main contributors to the model for the risk score.

Deployment

Deployment

Hard to trust Artificial Intelligence?

- Keep the model in a testing phase for at least 6 months.
- Keep comparing the risk score assigned by the experts to the model's prediction.
- If the margin of error is +-1%, deploy the model (any cloud solution or your own server).
- Keep coming back to the model every few months to see how it is performing.

You can always combine the risk score given by the experts with the model outcome using a weighted average.

Future Avenues

Future Avenues

- We could see that the model performs extremely well only with the numerical and categorical data.
- You can add textual as well as image data captured from the applicant's documents to improve the model. This would open up avenues to use strong Natural Language Processing (NLP) models like BERT to improve the results.

Thank You,
from thinktum

ACTIVATE
VIRTUAL SEARCH & AI CONFERENCE
Powered by Lucidworks