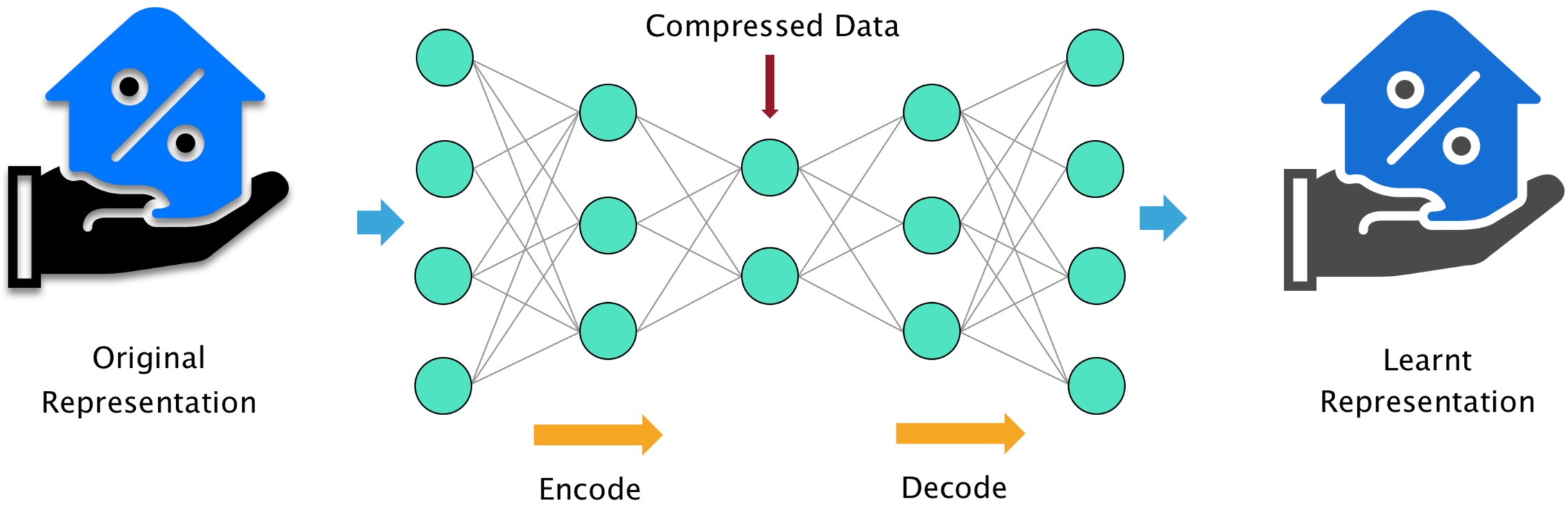
Applications of Deep Learning: Credit Risk Analysis

Student: Mike J. Cooper, Supervisor: Dr. John Cartlidge, Project Type: Research

University of Bristol, Department of Computer Science

Figure: A representation of an Autoencoder



Motivation

In recent years, deep learning has become a very popular machine learning method. This has mainly been due to three aspects, increased availability of high quality data, improvements in computational processing capacity (GPUs specifically) and more easily accessible frameworks such as Google's Tensorflow.
Unfortunately, the financial industry is generally slower at adopting new technologies. However, it is also an industry where high quality data is widely accessible and is one reason why Financial Technology (FinTech) industry is growing very quickly.
There was an opportunity to work with an industry partner who has made a large propriety data set available. The dataset comprises of over 120 million mortgages originated across the US between 1999 and 2016. There has previously only been one published paper [1] in 2016 on a dataset of equivalent size in the domain of deep learning. However, there has been extensive research published on much smaller datasets. Interestingly, there are a number of techniques used that were not applied to the 2016 paper. The goal of this research is to evaluate whether these same improvement method are successful when applied to a larger dataset.

Data

1. **Dataset A** - German credit (Size: 1000)
2. **Dataset B** - American Credit (Size: 130 million)

Project Objectives

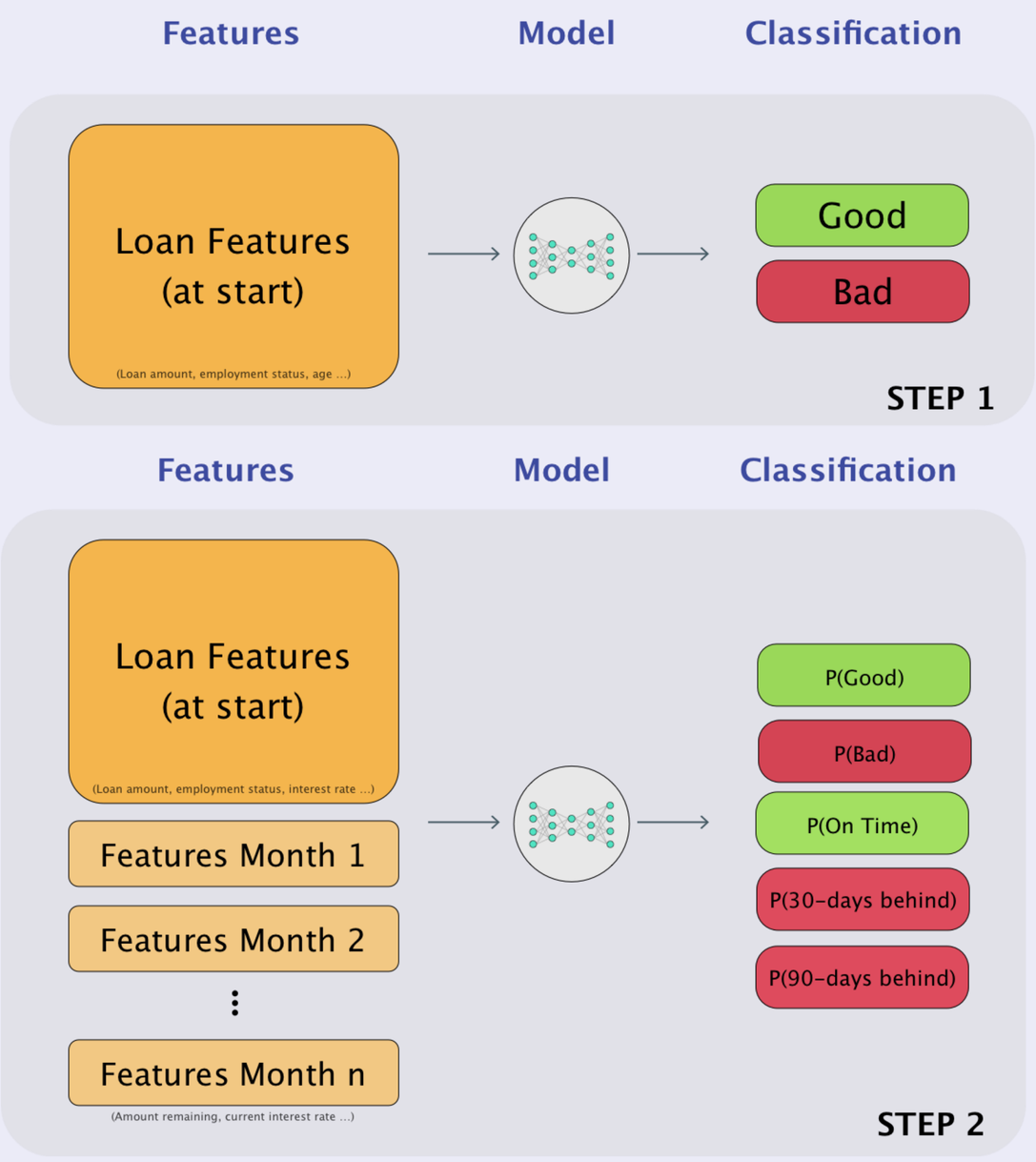
This project has the following objectives:

1. Reproduce research results using Dataset A
2. Evaluate if improvements from existing research can be applied to subset of Dataset A.
3. Introduce monthly features to Dataset B
4. Introduce other data to improve classification (e.g. US interest rates, housing prices)
5. Classify loans using time-related forecasting e.g. p(loan paid on time in a year from now)
6. (extra) Experiment with using a Conventional Neutral Network

Project Status

1. Replicated results from [4] [3] [2] using Dataset A
2. Data-processing and pipeline for Dataset B completed
3. Thesis paper structure formalised and created.

Classification of Loans



References

[1] Sirignano, Justin and Sadhwani, Apaar and Giesecke, Kay, Deep Learning for Mortgage Risk

[2] A. I. Marqusa V. Garcab and J. S. Snchez, Exploring the behaviour of base classifiers in credit scoring ensembles

[3] Van-Sang Ha, Ha-Nam Nguyen, Credit scoring with a feature selection approach based deep learning

[4] Ala'raj, Maher, and Maysam F. Abbod. Classifiers consensus system approach for credit scoring