Case Study: Using Artificial Intelligence to Improve Underwriting

underwrite.ai

Credit Underwriting Today

- Created in the late 1950s by Bill Fair and Earl Isaacs at the Stanford Research Institute and later at Fair Isaacs.
- Typically based on linear regression.
- Creates scorecards (rulesets) based on four key attributes:
  - **FICO score** (35% Payment History; 30% Revolving Utilization; 15% Credit History Length; 10% Types of Credit Used; 10% Inquiry Count
  - **DTI**
  - **Trade line performance** (e.g. 30 day late payments)
  - **Inquiry count**

CAN YOU SEE THE PROBLEM?

Why This (Kinda) Works

- As long as you limit lending to the Prime market, people with credit scores at or above 680, then it works pretty well
- You've screened out the population most likely to default
- What's left may not be optimal, but it's generally good enough, if your cost of funds is low

What's Wrong With This Approach?

- But what about the 51% of the population below 680?
- How about the 35% of the population below 620?
- Or the 25% of the population with no score?

The current models do a very poor job of predicting risk in these populations.

Results Of This

- Consumer lending is largely restricted to people with a history of consumer borrowing
- Because models of subprime lending risk are inherently flawed, lenders can only make a profit by inflating the overall lending rates

Controversial Statement

Essentially, prime lending is subsidized by the overcharging of subprime customers.

Current Status

We've divided the population into three classes.

- People with access to cheap credit
- People with access to expensive credit
- People with access to no credit

Another Controversial Statement

The problem can’t be fixed with the current underwriting methodology, because four attribute classes aren’t enough to model the problem.

Lenders are trying to solve a nonlinear problem with a linear solution. The linear solution works for that portion of the problem that can be solved in a linear fashion, but that only addresses one half the population.

So Why Not Just Add More Attributes?

- If we limit the universe to people with FICO scores between 550 and 850, and DTI below 40%, and late trades no more than 5 in each of three time ranges, and inquiries to 10 within 6 months…
- We still have 15,750,000 possible combinations of attribute values
- Representing this many possibilities in a scorecard requires a significant number of rules implemented in code, even using coarse tiering

So Why Not Just Add More Attributes?

- Adding another score with 700 values, increases the range of possibilities to 11,025,000,000
- This is why when you tell your programmers to “Just add another attribute - how hard can that be?”, they look at the floor

What Happens If You Want To Use 25,000 Attributes?

- That was the question that I asked about 3 years ago
- I wanted to understand the correlations between patterns of gene expression and prostate cancer when we had expression values for 25,000 genes

Machine Learning Provides the Answer

- Ensembles of machine learning algorithms allow us to efficiently model the expression patterns of 25,000 genes.
- 2,400 attributes taken from a subprime consumer credit bureau is a piece of cake!

What Came Next?

I set out to build an online system that would take application attributes and a credit bureau file and apply machine learning algorithms in real-time to return a probability of a binary outcome, e.g. charge-off vs full repayment, profitable vs unprofitable, positive lifetime value vs negative lifetime value.

Client #1

- Our initial client was a deep subprime unsecured installment loan lender. Their basic product is a 5 month loan with 10 bi-weekly payments.
- When we got involved their First Payment Default rate was 32.8%.

First Payment Default (FPD)

- This means that 32.8% of loans failed to make the first scheduled repayment
- Surprisingly enough, this is not uncommon in their industry sector

First Months Outcome

- By building a non-linear model set using our ensembles, we reduced FPD from 32.8% to 22% in the first month

Second Month's Outcome

- By feeding the results of the first month's lending back into the model and retraining, the FPD went down in month two from 22% to 18%

Month Three

- By continuously updating the model with new performance data we hit 15% in month three

Months 4-6

- Month 4: 12%
- Month 5: 10%
- Month 6: 8.6%

Subsequent Months

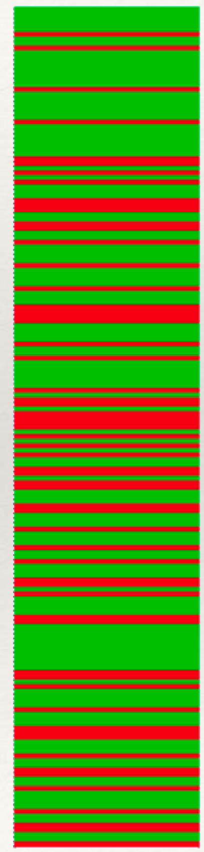
- We were able to determine that 9% FPD was the sweet spot for this lender. Further reductions produced diminishing returns. So we slightly altered the model to go from 8.6% FPD to 9% FPD
- This model has been running for two years.
- Not only is FPD still at 9%, but overall default is under 14%.

What This Does

This allows the lender to loan more money at better rates, to a larger population.

So, How Does this Really Work?

If we look at the 32.8% FPD distribution, where Red is FPD (Bad) and Green is Successful First Payment (Good), then we see the following time based distribution.



So, How Does this Really Work?

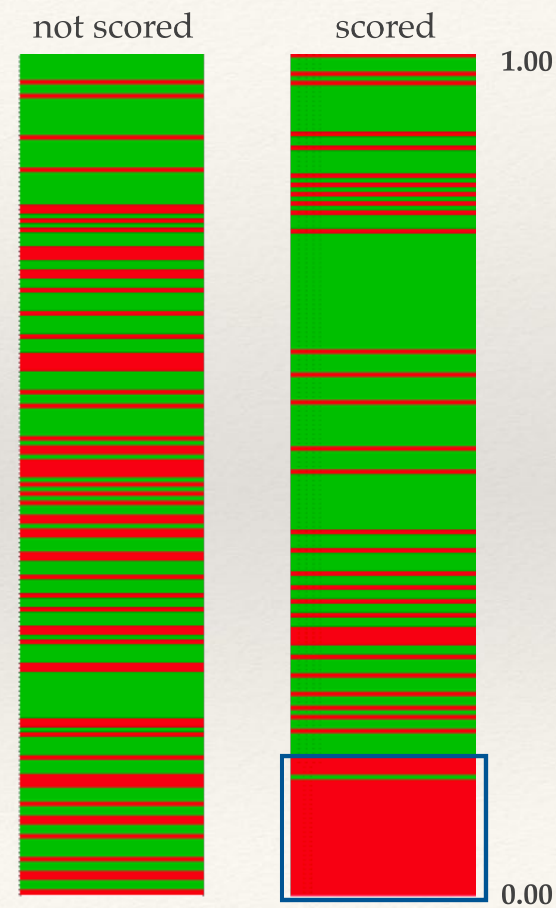
In our first pass of scoring these loans and ordering by our scoring model, here’s what we saw.

Scores range from 0 to 1 with 0 representing the lowest probability of a good outcome. Notice the density of Red (Bad) loans at the low end of the scale.

This type of clustering is what we are looking for.

So, How Does this Really Work?

We could identify 17% of the loans at the bottom of the scale that could be eliminated with negligible loss of good volume. This makes the balance of the loans markedly more profitable.



What Makes this AI?

- What makes this artificial intelligence is that the system we wrote **learns by itself**.
- We don’t “update the rules”. The model learns over time. It is fully dynamic. A key benefit of this, is that, changes in macro-economic conditions are reflected in changing outcome probabilities. Because the system is learning in real-time, lenders aren’t exposed to performance gaps caused by market changes.

What About Other Types of Lending?

I applied the same methodology to an analysis of corporate bankruptcy in France using standard financial ratios, and here's what I saw.

38% of corporate bankruptcies could be predicted with 100% accuracy and 45% with 94% accuracy.

What About Other Types of Lending?

We see similar performance gains across a range of lending areas, including:

- SMB lending
- Peer to Peer marketplace assets
- Corporate lending to public companies

What's the Secret Sauce?

- Replace linear regression based models with non-linear ones
- Expand the range of input variables exponentially
- Continuously retrain models in a production environment
- Understand quantum effect in consumer lending and the collapse of outcome uncertainly.

Questions?