I. Credit Scoring

II. Frequency-Severity Model

Dr. Rawichote Chalodhorn, Data Scientist, MIDAS

Credit Scoring

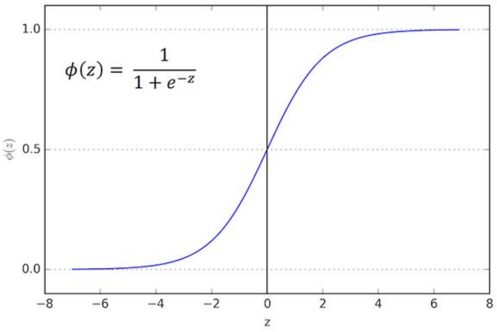
Lenders, such as banks and credit card companies, use credit scores to evaluate the potential risk posed by lending money to consumers and to *mitigate losses due to bad debt*. Lenders use credit scores to determine:

- Who qualifies for a loan?
- What interest rate should be for a borrower?
- What credit limit should be for a borrower?

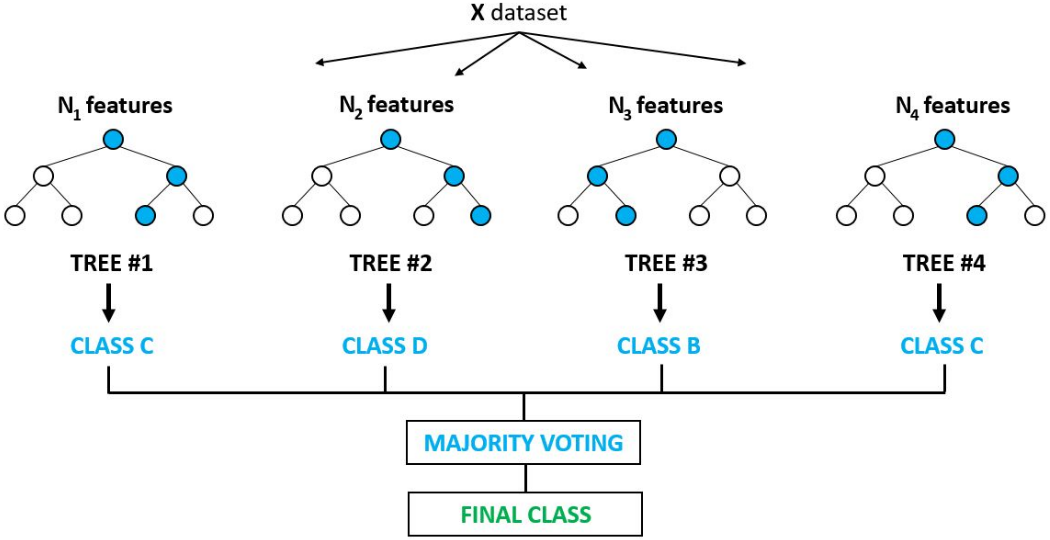
Typical ML Methods for Credit Scoring

Logistic Regression

$$p = \frac{\exp(\beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_n \cdot x_n)}{1 + \exp(\beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_n \cdot x_n)}$$



Random Forest



Credit Scoring: Deep Learning

ELSEVIER

Engineering Applications of Artificial Intelligence

Volume 65, October 2017, Pages 465-470

A deep learning approach for credit scoring using credit default swaps

Cuicui Luo, Desheng Wu, Dexiang Wu

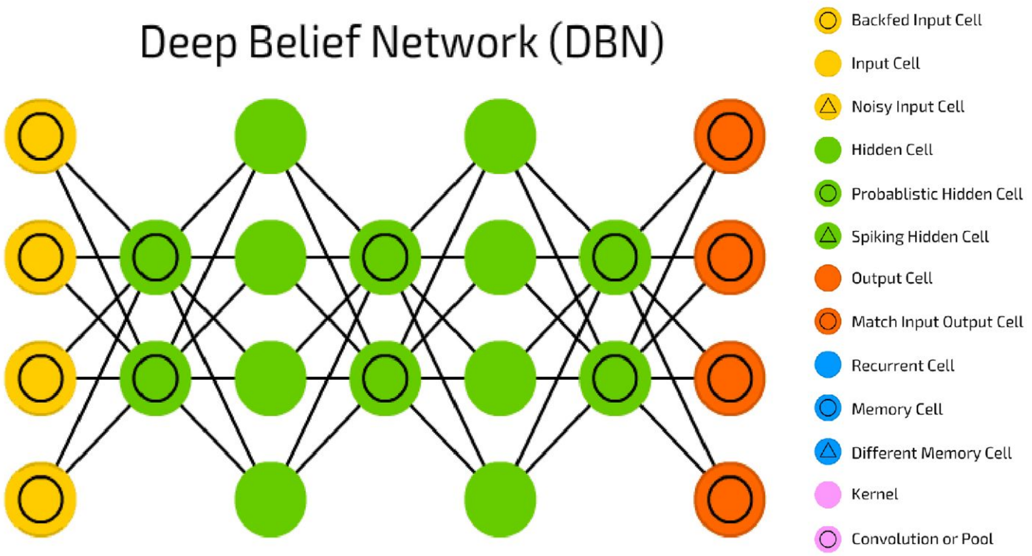
⊞ Show more

https://doi.org/10.1016/j.engappai.2016.12.002

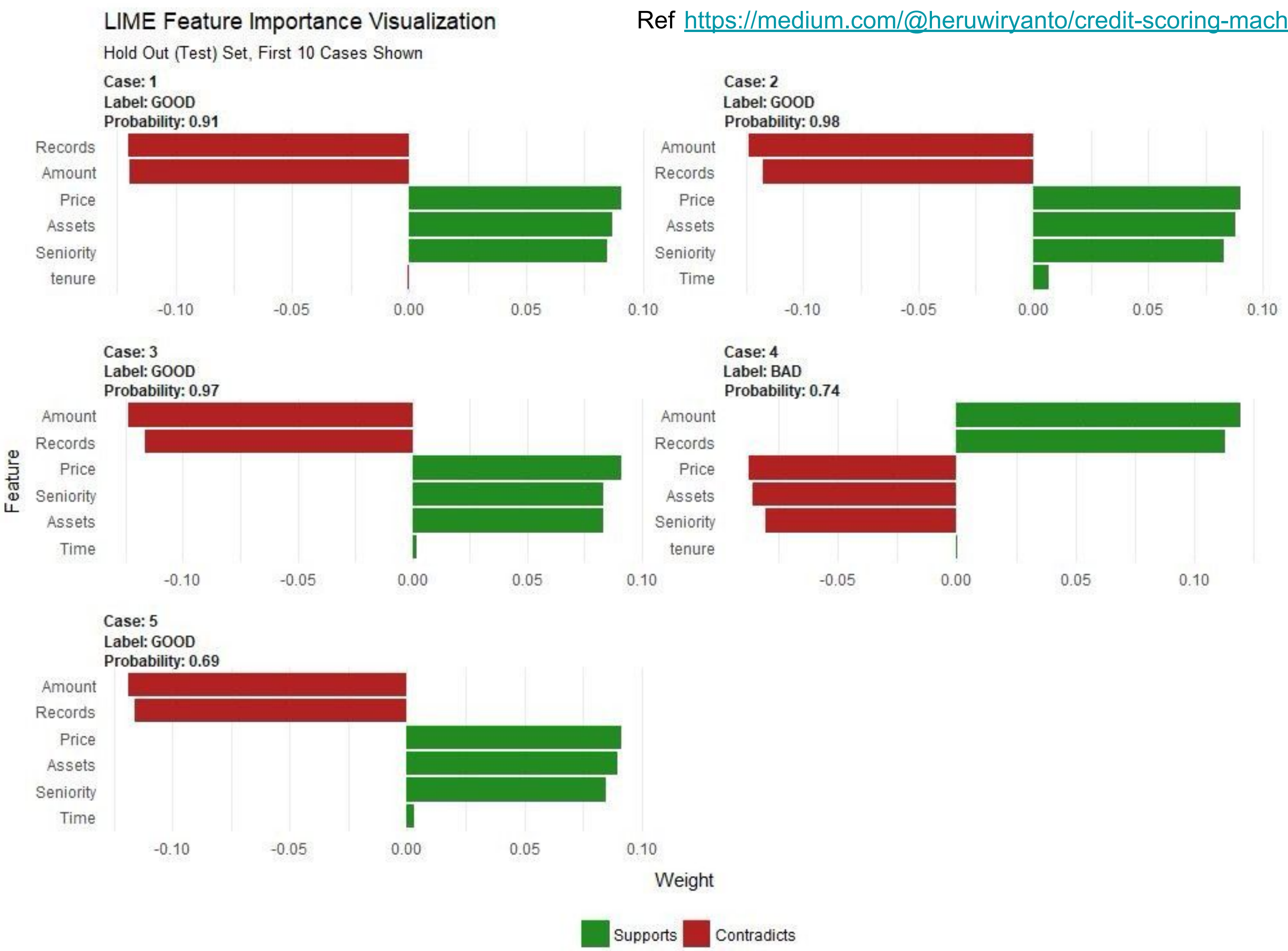
"... This paper investigates the performances of credit scoring models applied to CDS data sets. The classification performance of deep learning algorithm such as deep belief networks with Restricted Boltzmann Machines are evaluated and compared with some popular credit scoring models such as ***logistic regression****,* ***multi-layer perceptron*** *and* ***support vector machine****. The performance is assessed using the classification accuracy and the area under the receiver operating characteristic curve. It is found that* ***DBN yields the best performance****."*

Credit Scoring: Deep Learning

- DBN can learn to probabilistically reconstruct its inputs.
- The bottleneck layers then act as feature detectors.
- After input-training step, a DBN can be further trained with supervision to perform classification.



Ref https://medium.com/@heruwiryanto/credit-scoring-machine-learning-with-keras-r-502fc6eb451d



Features

1 Status	credit status
2 Seniority	job seniority (years)
3 Home	type of home ownership
4 Time	time of requested loan
5 Age	client's age
6 Marital	marital status
7 Records	existance of records
8 Job	type of job
9 Expenses	amount of expenses
10 Income	amount of income
11 Assets	amount of assets
12 Debt	amount of debt
13 Amount	amount requested of loan
14 Price	price of good

Credit Scoring by Using Phone Records

International Conference on User Modeling, Adaptation, and Personalization
UMAP 2015: User Modeling, Adaptation and Personalization pp 195-207 | Cite as

MobiScore: Towards Universal Credit Scoring from Mobile Phone Data

Authors | Authors and affiliations

Jose San Pedro, Davide Proserpio, Nuria Oliver

Conference paper
First Online: 11 June 2015

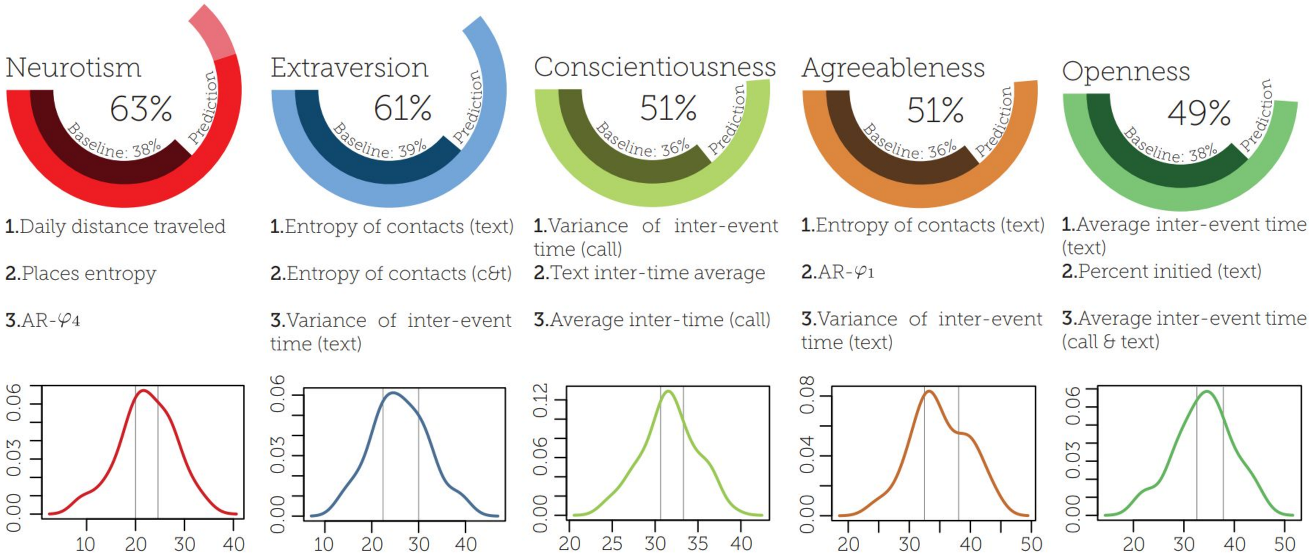
5 Citations | 1 Mentions | 12 Readers | 1.6k Downloads

- Research done in large scale in Latin America
- For people who don’t have bank records
- Estimate personality traits or socioeconomic status
- Use SVM

Predicting Personality from Mobile Phone Data

International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction
SBP 2013: Social Computing, Behavioral-Cultural Modeling and Prediction pp 48-55 | Cite as

Predicting Personality Using Novel Mobile Phone-Based Metrics



Frequency-Severity Method

Frequency-severity method is an actuarial method for determining the expected number of claims that an insurer will receive during a given time period and how much the average claim will cost. *Frequency-severity method uses historical data* to estimate the average number of claims and the average cost of each claim. The method multiplies the average number of claims by the average cost of a claim.

$$\frac{\text{loss amount}}{\text{premium}} = \underbrace{\frac{\text{loss amount}}{\text{claim count}}}_{\textbf{Severity}} \times \underbrace{\frac{\text{claim count}}{\text{premium}}}_{\textbf{Frequency}}$$

Tweedie distribution

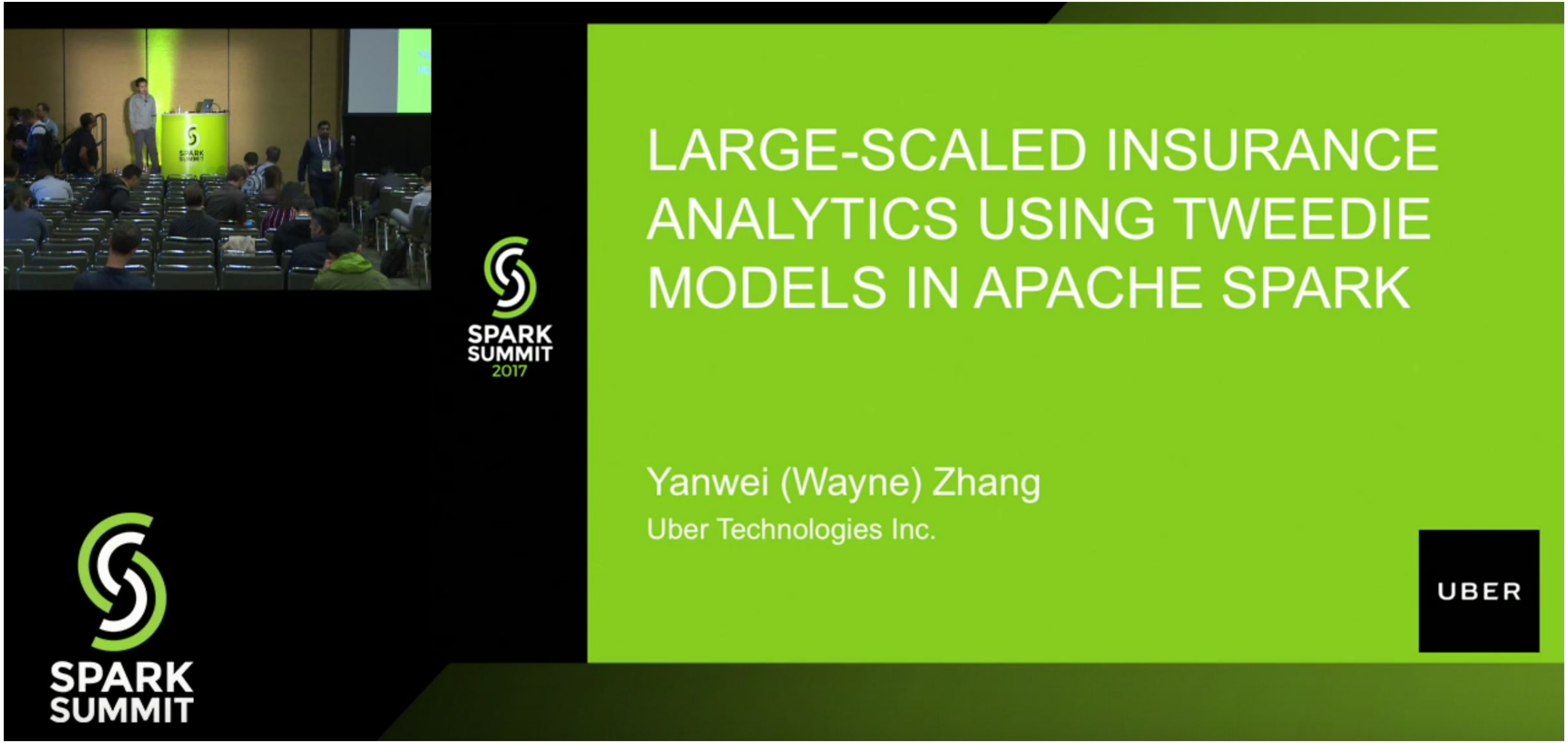
Generally, a logistic model or a Poisson generalized linear model (GLM) is appropriate for frequency, and a gamma GLM works for severity.

Recently, actuaries have modeled loss costs directly using the Tweedie distribution.

What is a Tweedie distribution?

If claims occur through a Poisson process and each loss is gamma-distributed, then the total dollars of loss are Tweedie-distributed. That description makes it sound as if frequency and severity are independent. But if you dig into the equations, you'll find that's not the case; in fact, a **Tweedie GLM implicitly assumes that predictors of loss simultaneously increase or decrease both claim frequency and claim size**.

Tweedie Distribution Implementation



https://youtu.be/S4Y4SU1ExJA

Axa Large Loss Claim Prediction

	High Frequency	Low Frequency
High Severity	**Avoid**	**Transfer**
Low Severity	**Reduce**	**Retain**

Axa team turned to ***Deep Learning*** and developed an experimental neural-network model. They began ***by identifying 70 risk factors***; driver age, address, vehicle type, prior loss history, vehicle age, original purchase channel etc. These 70 factors were then entered into a fully connected neural network with three hidden layers. This approach ***achieved 78% accuracy in its predictions*** over the ***40% baseline from a random forest model***.

Feature Extraction from Unstructured Data

Image recognition by convolutional neural network

Rice field : 95%
Flooded : 20%
Drought : 5%

Rice field : 90%
Flooded : 85%
Drought : 2%

Rice field : 65%
Flooded : 0%
Drought : 93%

Feature Extraction from Unstructured Data

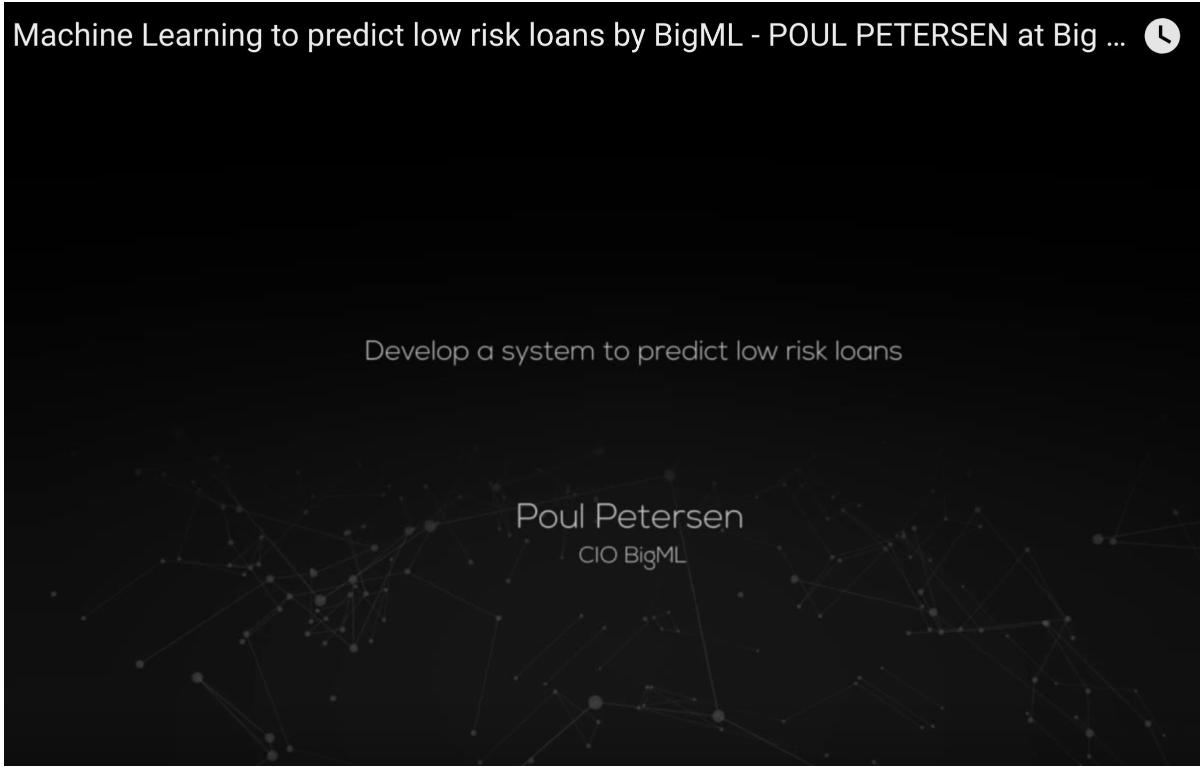
http://precisionagricultu.re/satellite-imagery-as-a-data-source-for-prescription-and-precision-farming-in-australia/

Satellite Image

utc_datetime	lat	lng	radius	image_no	est_flood	est_drought
2004-09-17T00:00:30.75	14.831440	99.939580	100	00-000-00001	0	0
2004-10-17T00:00:30.76	15.831440	99.909670	80	00-000-00002	1	0
2004-12-17T00:00:30.77	13.831640	99.939580	200	00-000-00003	2	0
2005-01-17T00:00:30.78	14.631340	101.438058	300	00-000-00004	3	0
2005-03-17T00:00:30.79	12.232440	100.134558	30	00-000-00005	0	1
2005-04-17T00:00:30.85	16.220144	99.939580	150	00-000-00006	0	2
2005-09-17T00:00:30.81	17.921040	99.939580	170	00-000-00007	0	3

Auxiliary features

Bonus Slide for Q&A



https://youtu.be/wvhqcMse70c

This YouTube video could answer many questions that I got in the Q&A session. … So, enjoy. ;-)