Eric Jenvey

Leader | Data Scientist | Builder

My Career Journey

2008 - 2012
Studied **Systems Engineering** at University of Virginia

2011
My **Data Science** journey began at Elder Research

2012 - 2018
Data-focused consultant at CapTech (Data Science, BI, Data Engineering, and more)

2018 - 2022
Founded the data science and analytics team at McGuireWoods, an AmLaw 50 Law Firm

2022 and beyond
Your great company?!

Predicting Settlements

The problem

*One of the most **significant** projects from my time at McGuireWoods.*

CONTEXT | We represented a **Fortune 50 logistics company** for all of their **truck accident litigation** (1000s of cases annually).

Our attorneys and the client would use data in a **descriptive, reactive** manner to adjust their strategy quarterly/annually, but had **no method of using the data proactively**.

ISSUE | Without using this data to quickly assess the riskiness of cases, we were unable to:

- **Properly allocate experienced resources** to riskier (i.e., higher settlement) cases
- Give the client an **accurate estimation of total risk** portfolio across all litigation
- Set **mutually agreeable terms** with opposing counsel, instead giving an opening to the opposition to set the terms

Predicting Settlements

The solution

		PREDICTED classification			
	Classes	a	b	c	d
ACTUAL classification	a	TN	FP	TN	TN
	b	FN	TP	FN	FN
	c	TN	FP	TN	TN
	d	TN	FP	TN	TN

APPROACH |

- **Exploratory data analysis** to gain trust and propose an ML approach.
- Surveyed the attorneys for expert hypotheses surrounding **why a case may settle higher or lower**.
- Constructed features from our data, such as **injury severity**.
- Tested various models to predict settlement, and after tuning landed on a model with **Mean Absolute Error of around $50K**.
- Unfortunately, this error was too large for the case portfolio, and the model would lack credibility if deployed widely.
- In response, we changed our approach to a **classification model**, which we termed as “bucketing risk” into High/Medium/Low bands.

SUCCESS | Even though we lost some information by simplifying our prediction, the model became much more accurate, and resulted in **significant risk reduction** for the client. We were able to use the predicted settlement as a major component of case strategy, and change the paradigm of our client representation from **reactive to proactive**.

Cutting Over-Budget Cases in Half

*I am really **proud** of this project - we successfully understood our customer's needs and knocked it out of the park. We were even **recognized by the Financial Times** for the solution!*

The problem

CONTEXT | Law firms **traditionally bill clients by the hour**. So, the more you bill, the more $$ you generate. The firm had arrangements with multiple major lenders to handle **single plaintiff disputes** on a **fixed fee basis**. Therefore, the **more we bill, the less cost-efficient our time**.

ISSUE | My analysis at the outset of the project indicated that, historically, **75% of cases** were "underwater" (i.e., more hours billed to the case than what the fixed fee would support). It was evident that a **major part of this issue was an information gap**. Our attorneys needed data in their hands to make decisions about staffing, case strategy, etc.

CHALLENGE | Our primary challenge was to develop an **all-in-one dashboard** to support **both** individual contributors (i.e., **tactical users**) and managers (i.e., **strategic users**), which could account for case performance across a number of dimensions. This dashboard should also be **tailored to each user,** so that an attorney was able to assess performance for their specific cases, rather than ones irrelevant to his/her portfolio.

We also needed to **replace and retire the existing reporting**, which consisted of weekly emails generated through SSRS, only sent to a select group, and not tailored to highlight cases with major issues, as there was not a unified KPI.

Cutting Over-Budget Cases in Half

The solution

APPROACH |

- Held **design sessions** with our stakeholders, the Partners, to understand their requirements.
- Developed a **data architecture** to hold both actual hours/$$ (ETL'd from our source accounting system) and budget (previously tracked manually).
- Created a **dashboard** targeted to both individual contributors and managers on the portfolio, specifically tailored to the cases relevant to each user.
- Ran a **pilot program** and experimented with features to further **refine** the dashboard.

SUCCESS | Due to the massive increase in availability and freshness of the data, our attorneys had much greater control over matter staffing, leverage, and approach, and over budget cases were significantly reduced **from 75% to around 35%.**

Task Code Classification

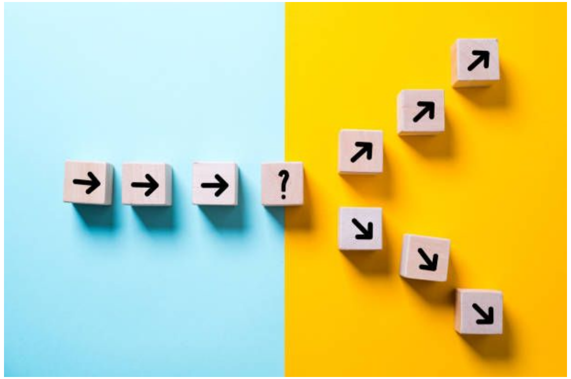
The problem

Sharing one of my ***greatest learning experiences.*** *This was a failed launch that taught me many good lessons that informed the remainder of my work at McGuireWoods.*

CONTEXT | The core transaction at a law firm is **time entry**. Attorneys perform work on behalf of their clients and record that time. The time entry kicks off a number of processes that result in the firm being paid for that time. These time entries, and their associated descriptions ("**narratives**") provide an incredibly detailed window into the work being done at a firm.

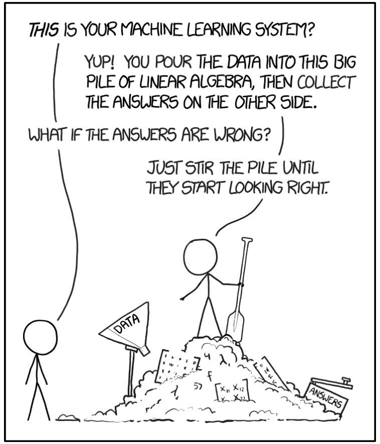
ISSUE | Most law firms do not spend time combing through this data. Instead, they rely **solely on the individual, subjective experiences of their lawyers** to make decisions around:

- Optimal case staffing
- Promotions + separations
- Hot sectors and associated staffing
- Markets the firm should pursue further
- Concentration of attorney expertise



Task Code Classification

The solution & lessons learned



APPROACH |

- Objective was to **turn these time entries into a rich dataset** to empower firm leadership to truly understand and answer the questions posed on the last slide
- Developed a **common set of task codes** to classify attorney work based on client/context
- Our attorneys and data entry team hand-labeled a large sample of time records with task codes.
- **Applied NLP techniques** to clean the text, tried different word embedding techniques (landed on **TF-IDF scores**)
- Tested a host of classification models. Our best performing models **classified roughly 80% of our time entries correctly**, a great result!

TAKE AWAYS | We could not reduce the runtime of the model to an acceptable threshold for our time entry system, and thus could not deploy it in an online manner (the business wanted attorneys to confirm or decline the task code while entering time). But among the failures, I learned a **few good lessons**:

- **Batch models** can still be valuable to the business
- Importance of **balancing** accuracy/efficacy with computational efficiency
- Cross-team resource **planning** and project **prioritization**
- The immense **value of ML** engineering!

Eric Jenvey