Artificial Intelligence for Text: Deadly Insights and NACE

James Kelly

Why Text Classification?

Survey Data
E.g. Census, SILC

Administrative data
E.g. PMOD† from Revenue, PPOD* from Dept. Education

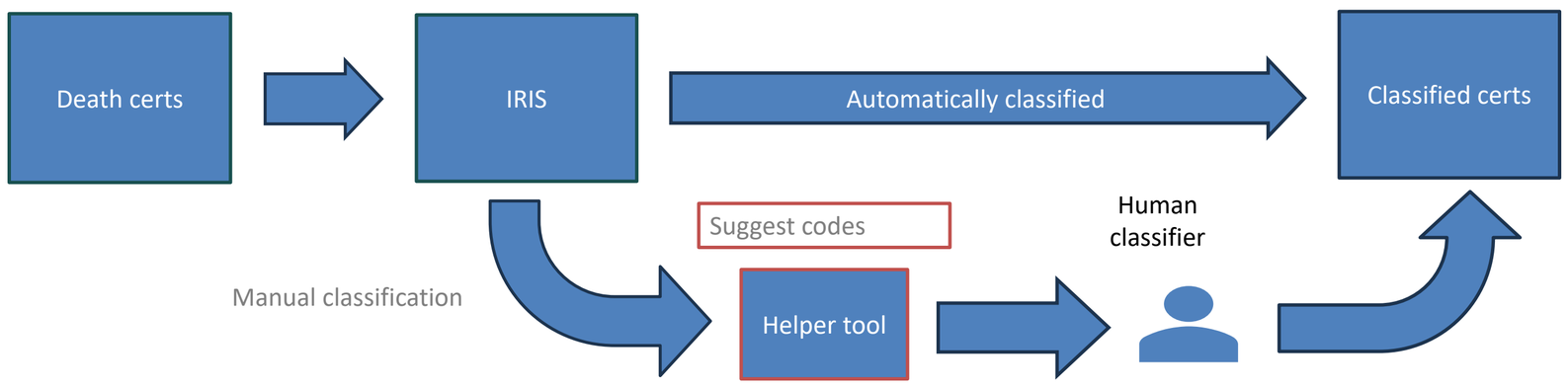
Novel data sources
E.g., WIN and job advert, web-scraped data, company annual reports

Increasing opportunities with free-text data

\+ Improved timeliness, coverage (and granularity), lower response burden

\- Less control over structure of data, systematic issues with availability and accuracy

Cause of Death Classification

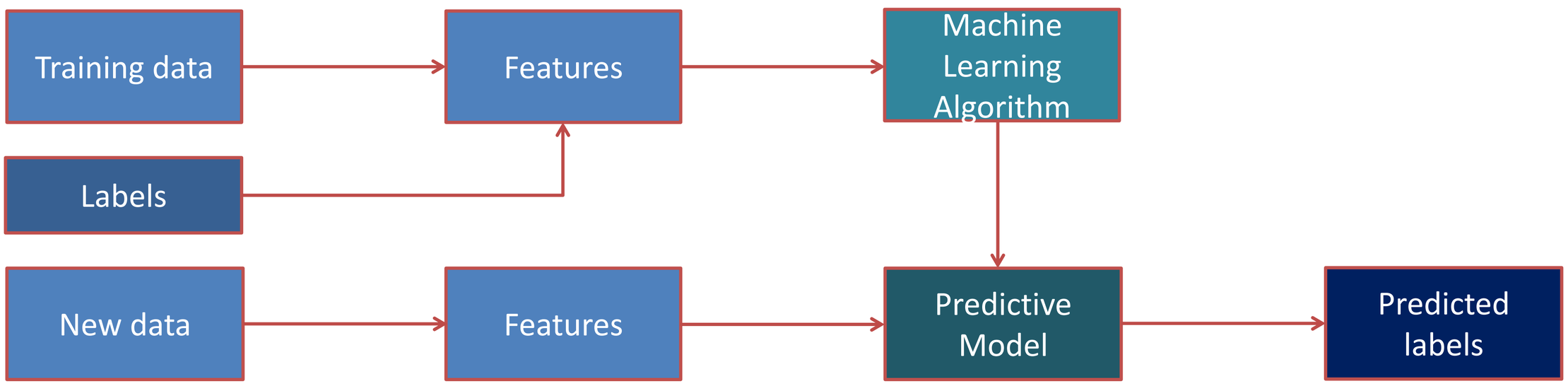


The Vital Statistics division of the CSO must classify 35,000 deaths each year for statistical output.

Approximately 50% are automatically coded, but the rest are manually coded.

An A.I. solution that aids the manual coders was identified as one way to greatly improve this coding process.

How does Text Classification work?



Final Product

I designed and implemented a shiny app as an interface between the manual coders and the final machine learning model.

The app allows the coders to connect to the certificate database. The tool then breaks each line into its component parts and gives the top 5 most likely ICD-10 codes for that text.

The Coder can select a code which will link them to the official WHO ICD-10 website. Using their domain expertise, they can then make a robust classification decision.