Our Vision

Despite the enormous recent progress in AI and machine learning (ML), the reality is that most organisations must deal with tabular data consisting of a mix of numeric and categorical attributes, often with highly skewed distributions and missing values. Such datasets continue to pose challenges, and it is in this space that we are focused.

At Skanalytix we develop methods and tools that allow organizations to better leverage their mixed-type data and enhance their data analytics capabilities by allowing insights not capable of being drawn using existing methods. Whether your focus is on data security and intrusion detection, data privacy, developing a recommender system, or countless other data-centric tasks, if you deal with complex mixed-type data then there is a place in your organisation for our unique data modeling and inferencing framework.
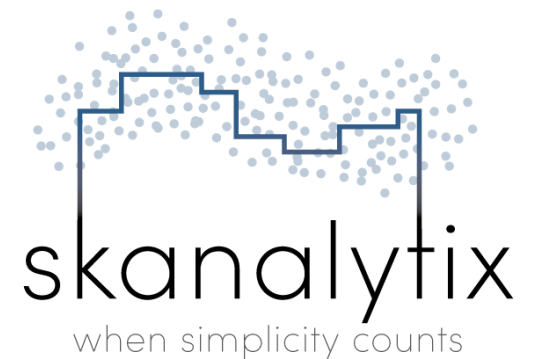
Find out more by visiting our website or contacting us directly to explore how our services can be used to complement your existing data analytics capabilities.

Contact Us

skanalytix
when simplicity counts

A New Approach to ML Modeling

Given some dataset of interest, a typical ML task might go something like this: (1) choose the variable you wish to predict; (2) select the type of model you wish to use (decision tree, logistic regression, MLP, etc.); (3) optimize the parameters of the model; (4) use the model to make the prediction.

The problem with this approach is that if you then wish to perform a different task on the same dataset this whole procedure must be repeated. This is extremely inefficient.

At Skanalytix we went back to the drawing board – back to square one – and asked the question '*Is it possible to construct a SINGLE model from which a wide range of data analytics/ML tasks can be performed?*'

The answer is Yes, and the framework we have developed is **UNCRi – Unified Numeric/Categorical Representation and Inference**.

The UNCRi Framework

At the heart of the UNCRi framework lies a unique graph-based data representation scheme coupled with a powerful inference procedure. The representation scheme treats numerical and categorical variables in a unified fashion and means that a single robust measure can be used to calculate the similarity between any two datapoints, regardless of the mix of variable types or distribution of values. However, these similarities aren't explicitly computed; instead, they are implicit in the graph-based inference procedure which can estimate the probability distribution of any target variable conditional on the values of one or more other variables.

The framework's versatility in estimating conditional distributions allows a range of generic data-oriented tasks to be performed. These include *prediction* (*classification* and *regression*), *imputation*, *joint probability estimation* and *synthetic data generation*. However, these out-of-the-box solutions only scratch the surface. The UNCRi framework's breadth and flexibility extend to custom tasks such as the development of recommender systems.

Application Scenarios

- *You are concerned about data security and want to develop a model for detecting intrusions.*
  Solution: Intrusions are often outliers in your system log data. Use UNCRi to estimate the joint probability of your log data and select a threshold below which you would consider a datapoint to be an outlier.
- *You need to share data for the purpose of conducting an analysis, but data privacy rules prevent this.*
  Solution: Use UNCRi to generate synthetic data points that are distinct from – but drawn from the same distribution as – the original data. Perform your analysis on the synthetic data.
- *You want to find clusters in your male, high-spending customers but you have insufficient datapoints.*
  Solution: Use UNCRi to generate synthetic datapoints matching these criteria. Append the synthetic data to the original data, and then perform clustering.
- *You want to create a recommender system.*
  Solution: The UNCRi framework can be used to create effective content-based, collaborative and hybrid recommender systems.

The UNCRi framework opens a world of possibilities!