PANDA^{TM}: Predictive Advanced Nonlinear Diagnostic Analyzer

PANDA is an advanced tool for predictive and analytic modeling. It is designed for use in a MATLAB environment. NOTE: PANDA ranked in top ten in the Heritage Health Provider Network Competition https://www.heritagehealthprize.com/, out of 1600 competitors world-wide. It offers several outstanding features.

1. **Inclusion of highly specialized learning models based on state-of-the-art optimization technology**
   The capability of any predictive and analytic modeling tool depends on the inclusion of accurate and diverse models in its underlying ensemble of learners.

   PANDA contains a wide collection of standard learners such as linear and nonlinear support vector machines, regularized logistic regression, bagged or boosted trees, and kernel kNN. More importantly, in addition to the standard learners, PANDA includes highly specialized learning models based on advanced optimization technology that can effectively handle nonlinear and interaction variable selection, highly unbalanced or rare class learning and semi- and unsupervised learning.

2. **Effective handling of nonlinear and interaction variable selection**
   While most off-the-shelf predictive and analytic modeling tools cannot learn complex interactions of input variables, PANDA can effectively handle the nonlinear and interaction variable selection by using advanced optimization technology to solve a complex mathematical problem that identifies the most important subset of variables.

3. **Rare class learning**
   Highly unbalanced datasets can introduce significant bias and present immense computational difficulties to the currently available predictive and analytic modeling tools. PANDA employs highly specialized rare-class learning algorithms that produce high-quality rankings with the ability to scale to very large datasets.

4. **Semi- and unsupervised learning**
   PANDA can take advantage of unbalanced examples to produce significantly better models.

PANDA can be applied to a variety of practical problems. One important area of application is computational finance, specifically, the predictive modeling of mortgage portfolios. In this application, PANDA^{TM} offers significant advantages over the currently available off-the-shelf predictive and analytic modeling tools.

1. **Proactive managing of the entire portfolio**
   PANDA can predict account status – current, delinquent, default, prepayment, and loss given default, and use model outputs to triage accounts, mitigate loss, and improve recovery.

2. **Advanced predictive modeling**
   Standard statistical 'scores' adjust slowly overtime and can be insensitive to changes in market conditions. Specifically, account status can deteriorate quickly in a volatile market. The account management needs much more timely responses than the existing tools can offer. To overcome this difficulty, PANDA employs advanced predictive modeling that makes use of the most data and gets the most out of the data to enable timely responses and actionable predictions.

3. **Use of comprehensive data**
   PANDA incorporates the following types of information -- loan type, interest rate, valuation, collateral, and outstanding principle. It can handle a variety of external variables such as credit bureau scores, Bureau of Labour statistics, and vendor data. It processes personal transaction-level data including category-based deposits and expenditures, and inflow or outflow by channel. It can also handle additional non-standard data sources such as social media websites.

About Cayuga Research

Cayuga Research is a consulting company focused on the development and implementation of advanced optimization methods. Our firm is composed of experienced research analysts and developers based in Waterloo, Ontario.