Closed Loop Continuous Learning with TIBCO

Case Study: Dynamic Online Pricing

Business Challenge

Dynamic pricing software has traditionally been dominated by a few specialist vendors that provide black-box point solutions. These solutions are slow to react to market trends, regulations, and the increasing volume and sophistication of underlying data. In the insurance space, companies are looking for ways to transform, differentiate, and compete. As quotes and premiums have moved online to self-service, systems for dynamic pricing have garnered much interest. There are some significant technical challenges however, including how to address the following issues:

- Insurance policy pricing is a sophisticated process, with multiple algorithmic components.
- Information required to generate technical and street premiums are often sourced from third-party data providers. How do you efficiently obtain this data?
- Acquisition of new customers and retention of current customers have their own algorithmic challenges. You want to preferentially capture and retain the best customers, those with the highest customer lifetime value, or you want to maintain a careful balance between supply of insurance capital, demand for insurance coverage, and unexpected external factors such as large catastrophes.
- Price optimization is its own science. While insurers want to maximize profits and grow the business, they also have to ward off increasing competition from both established firms and fintech startups.

There is a trend towards configurable online dynamic pricing solutions with closed loop and continuous learning. These solutions enable rapid reaction to market conditions via updates to pricing based on policy acceptance/rejection rates, sales campaigns, and/or digital marketing initiatives. In this whitepaper, we describe a TIBCO approach to such a dynamic pricing system. This solution includes configurations for price optimization and customer acquisition & retention, along with price elasticity data and modeling whereby acceptance and marketshare may be modeled as a function of price. The TIBCO solution is illustrated in the whitepaper with some elements of a solution developed and deployed at the Automobile Association of Ireland.

TIBCO Data Science, Machine Learning, and Visual Analytics

Price Optimization

Price optimization models have multiple components, including:

- Increase customer conversion and reduce customer churn rates
- Attract more new customers with competitive rates

Customer churn and acquisition models predict the probability of churn/acquisition for existing and new customers using historical quote acceptance information and a combination of customer demographics and vehicle attribution data (for motor insurance). Models are trained in TIBCO Data Science software as generalized linear regression models (GLM) with regularization via elastic net balancing between ridge regression and lasso methods. Coefficients for explanatory variables are constrained during the modeling to minimize the risk of violating civil rights and discrimination regulations.

Customer lifetime value, commission (profitability), and price elasticity (sensitivity to a price discount for individual customers) are used to segment the customer base and obtain optimal commission/discount tables.

To obtain churn or conversion probabilities, the GLM models are stored as PMML objects. These are used for real-time model inference against customer attribution data such as age, vehicle type, no claims discount, and customer membership duration (in years). Based on these estimated probabilities, optimal premiums and discounts are calculated for customer segments and automatically quoted upon customer application.

Acceptance and rejection data are obtained from each quote and stored, along with the quote details, in a relational database. As data accumulates, the TIBCO system triggers a model rebasing and promotion to production operations, with attendant internal approvals. This process can be scheduled with user-defined intervals (daily/weekly, etc.), or it can be activated based on events, such as conversion rate/log-loss ratio below a threshold and quote data accumulation. Model rebasing and promotion to production includes the following steps:

- *Model performance assessment.* This is based on a set of Gini coefficients. If outside allowable limits, models are re-evaluated automatically using a genetic algorithm for searching the parameter space. Candidate models are regularized with elastic net.
- *Model updates and explainability.* This includes assessment of business goals, variable importance, over-fit, bias or outliers, and regulatory perspectives. Segments and factors summarizing the differential model attributes (current versus prior model) are evaluated in a champion-challenger setting. Questions like “which customers are affected and how” are assessed.
- *Model diagnostics.* Assessment of local predictive power and accuracy across the predictor space is a consideration. Particular regions of the predictor space may be assessed in accordance with recent trends, and areas of concern can be identified. Model diagnostic metrics (AIC, BIC, ROC AUC), visualizations (ROC curves, lift charts), and statistical hypothesis test results are assessed.
- *Model versioning, approval, and audit*. In most cases, models need to be rebased on fresh data to account for variance in market conditions since the previous model training. As such, several versions of the models need to be saved and governed (per regulatory requirements). These models should also be available for auditing and compliance assessments. TIBCO Streaming software’s Artifact Management Server provides these features in an easy-to-use interface.

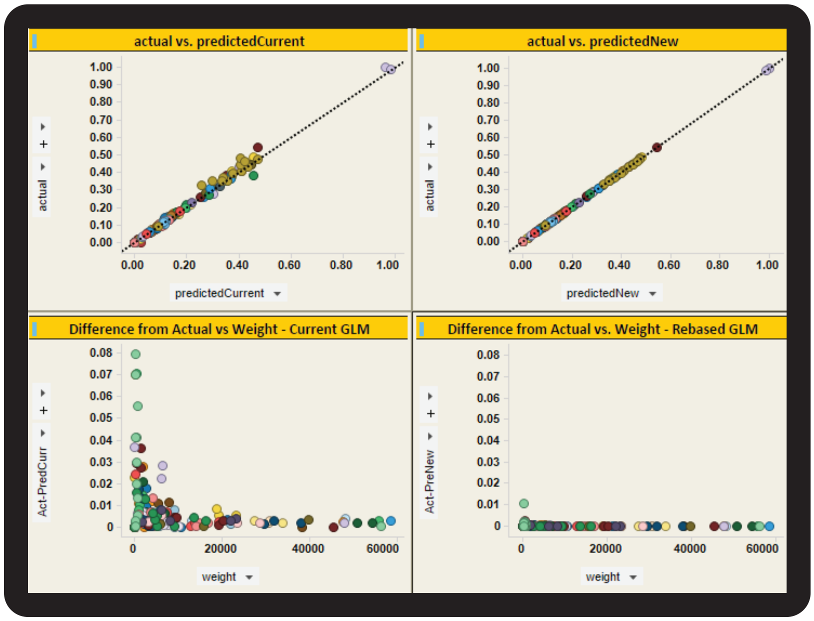


Figure 1a. TIBCO Spotfire dashboard components, model performance and diagnostics assessment for rebasing: Current and new rebased model.

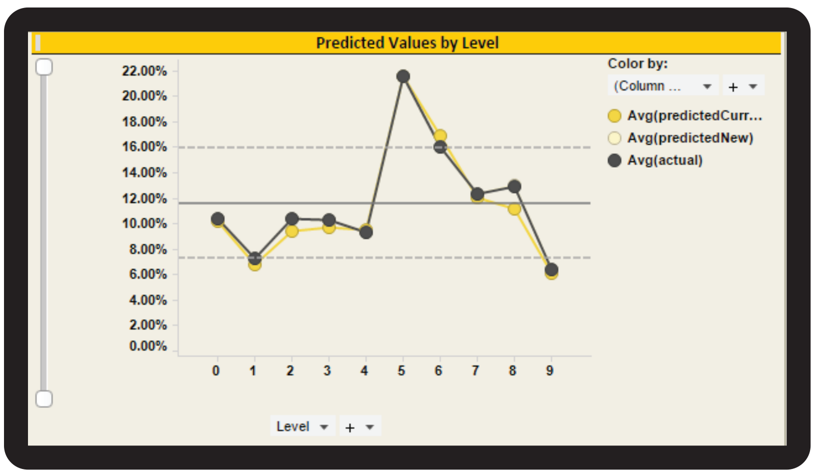


Figure 1b. TIBCO Spotfire dashboard components, model performance and diagnostics assessment for rebasing: Predictions for particular variables of interest to the business.

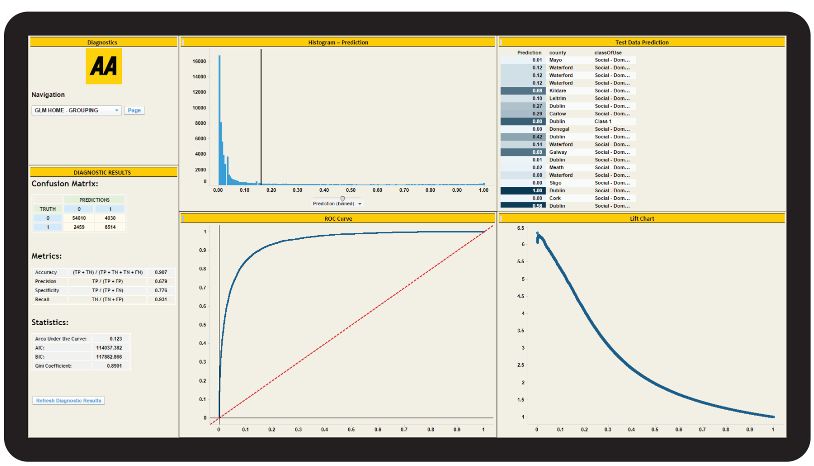


Figure 1c. TIBCO Spotfire dashboard components, model performance and diagnostics assessment for rebasing: ROC curves and confusion matrices.

Figure 2a. Model interpretability.

Figures 2a, 2b and 3 have been partially masked when it comes to the specific factors and variables. These figures are for illustrative purposes.

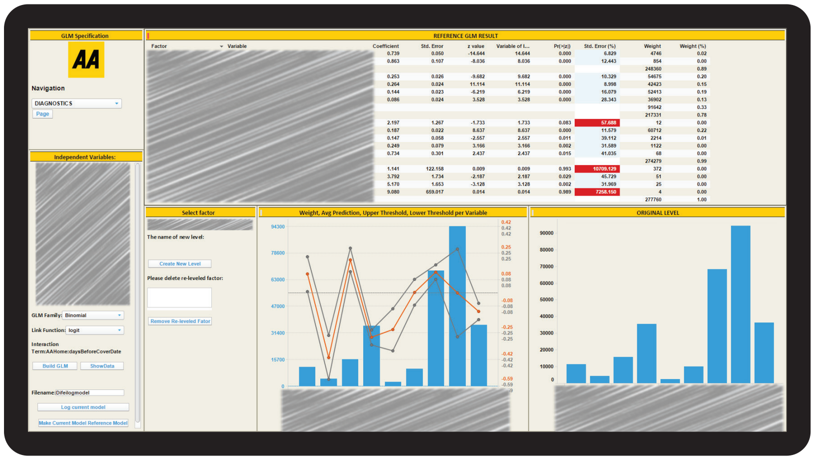


Figure 2b. Variable importance.

Customer Risk Scoring

Customer risk scoring models include functions to:

- Compute a customer quality score to help reduce fraud in policy quote requests and claims.
- Build models with recent data
- Run models in real time using actual customer inputs and consider price limits and fraud detection warnings
- Capture and detect potential manipulated quote requests

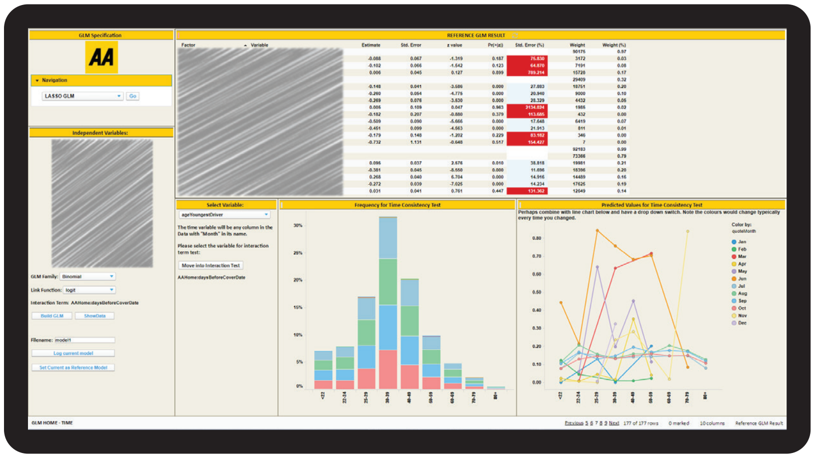


Figure 3. Spotfire customer propensity modeling: Predict customer propensity probability; compare true conversion rate distribution across time span and other predictors; update models in real time using actual customer inputs, and consider price limits and potential quote manipulation warnings.

Data science and machine learning models are built and deployed using TIBCO Data Science software, and presented using visual analytics in TIBCO Spotfire software.

The model management rebasing process is kicked off via a triggering mechanism that can be set based on a specific threshold or an external scheduler. This process can be scheduled with user-defined intervals (daily/weekly), or activated based on events, such as conversion rate/log-loss ratio and data accumulation thresholds.

The triggered job assembles the pertinent model performance, explainability and diagnostic data, and context, including customer risk, quote history, and demographics. TIBCO Streaming software kicks off the rebasing, and TIBCO Spotfire Automation software assembles the analytic dashboard. The heads of Pricing and IT review the dashboard, comment, and/or approve the model update to the production environment. The approval process includes provisions for detailed review, escalation, and resolution, though in most situations this process takes a matter of minutes or hours.

TIBCO Streaming Analytics

The Streaming Analytics component of the platform includes:

- Processing users' price/premium requests
- Modeling management and rebasing

Processing of user's price/premium request

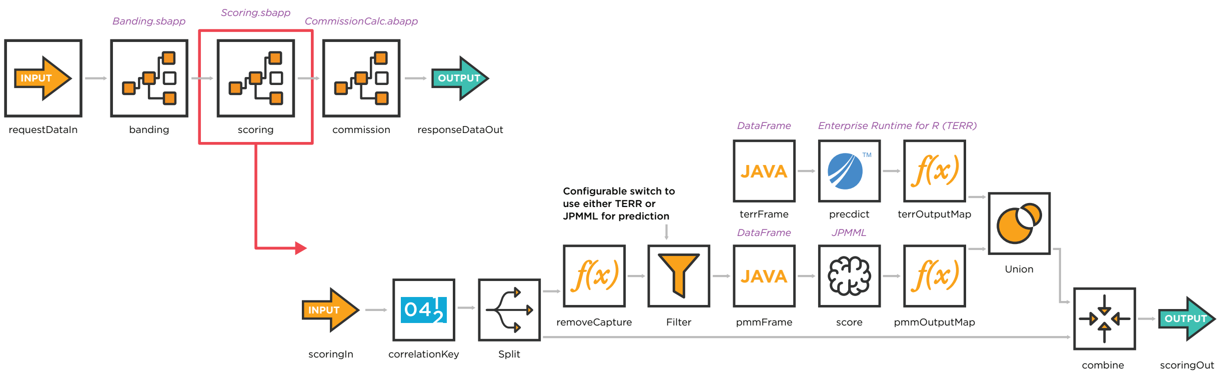
- Calculation of premium price and discount. The streaming analytics application aggregates the streaming data, performs the ETL process in real time, calculates the probability of a user's quote converting to a real sale, and calculates the corresponding commission and commission discount. This is achieved by scoring the current model in production, by passing the grouped factors and levels to the PMML adapter that scores the model on the fly. This end-end process takes less than 20 ms from user data entry to quote return to the user. This response time can be tuned to handle any reasonable input velocity, via available elastic scaling.
- Data and feature generation. To score the model, calculate the probabilities, and determine the commission and discount, the raw data that the user enters from the insurance company web portal must be converted into the features, format, and data types that are inputs to the hosted machine learning models. Based on predefined tables, a numeric variable age will be transformed into an age band, and factor levels are banded (grouped together according to occupation types). The banded values are then passed into the currently active GLM model.

Modeling management and rebasing

- *Model management.* GLM modeling objects are saved in a database table for compliance and auditing purposes. Alternatively, models/rules can be stored as artifacts within other TIBCO or customer applications.
- *Model versioning and updating.* Detects when a new model or commission table has been marked active in the data store, and then downloads and replaces its currently cached objects with the new active versions.
- *Model rebasing.* Assembles recent data (quotes and acceptance/rejection) blended with historical quotes. These are passed to a TIBCO Data Science workflow that re-trains the model on the accumulated data. After the updated training is complete, a TIBCO Spotfire Automation Services software job is triggered. This populates a TIBCO Spotfire model diagnostics and explainability template, and sends an email to appropriate business analysts. Finally, a model update approval process is initialized (for both pricing and IT functions).

- *Feature engineering and banding.* Analyzes quotes from a data warehouse to generate features/banding logic, and storing. The banded quotes are used by the GLM model's rebasing analysis in TIBCO Data Science software.

Figure 4. Processing the users' price/ premium requests, TIBCO Streaming event flow.

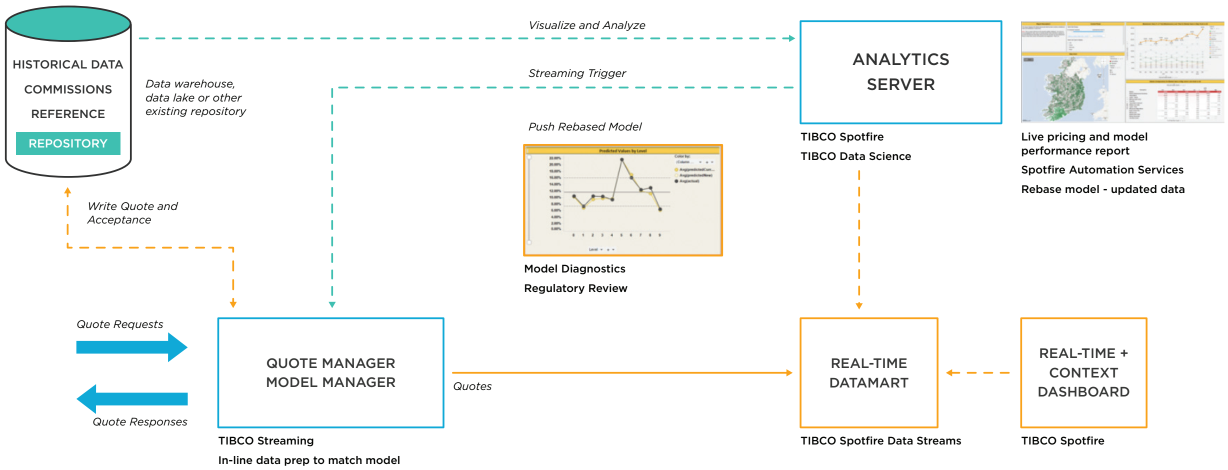


Solution Architecture And Components

The closed-loop, continuous learning architecture is outlined in Figure 5. Bottom left shows the quote manager request-response, which is served by the online dynamic pricing engine. This engine includes the current statistical pricing model (e.g. GLM or machine learning model such as GBM, random forest, etc.), which is rapidly scored on the incoming data stream (in less than 20 ms). Quotes and acceptance/rejection are written to the data warehouse (top left of diagram). As data accumulates, a model rebasing process is triggered. This includes model updates and automated dashboard creation with model summaries (for both champion, challenger model), explanations, and diagnostics. The web-based interactive model summary dashboard is automatically sent to stakeholders, including heads of analytics, pricing, and IT (top right of diagram).

If the rebased challenger model is accepted, it is pushed to the production quote manager system (middle of diagram). Quotes and acceptance/rejection information are also maintained in the TIBCO Spotfire Data Streams in-memory data mart, where real-time streaming data may be visually analyzed by interested parties, such as sales and marketing departments interested in the progress of various marketing campaigns in play.

Figure 5. TIBCO algorithmic insurer real-time pricing solution architecture.



To encourage interested users of the solution, the TIBCO solution assets are available from the TIBCO Community site. This includes all the scripts, configurations, and integrations required to run these solutions on other data. The TIBCO Community Exchange includes accelerators for various closed-loop continuous learning applications. The TIBCO **Pricing Accelerator** includes TIBCO Spotfire templates, TIBCO Data Science models, and TIBCO Streaming event flows.

Possible Extension

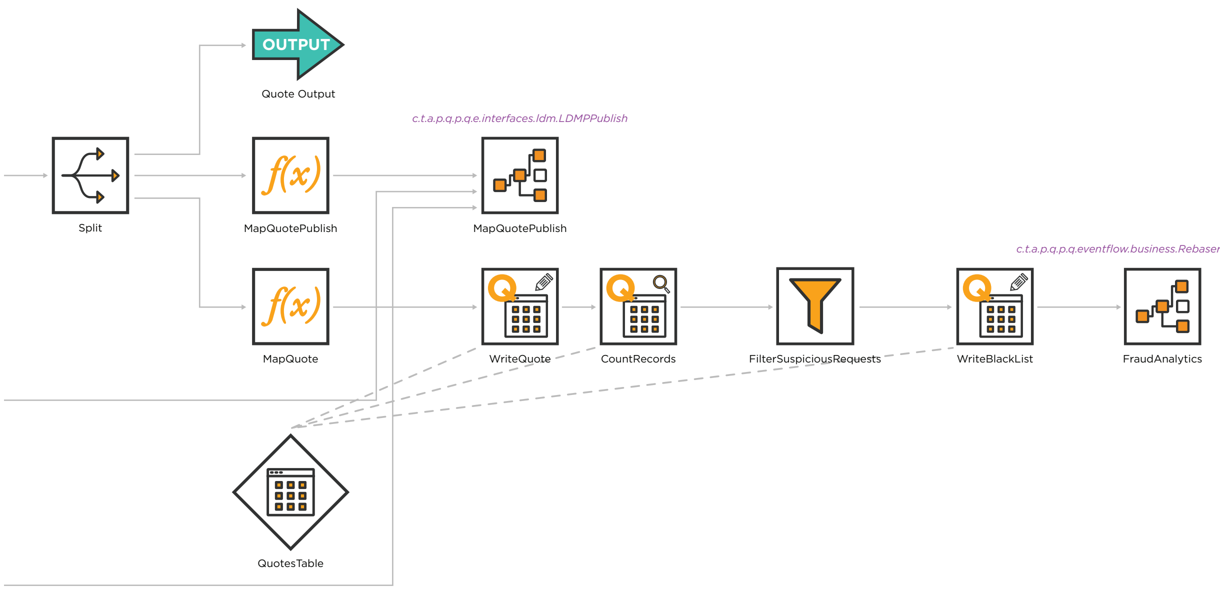
The generic nature of TIBCO solutions means the components of the platform can be modularized and re-deployed for multiple purposes.

For example, while a prospect is querying vehicle insurance, the same workflow can be extended to promote other products or services based on the result from market basket analysis or propensity modeling. In our example, vehicle insurance and home insurance tend to be popular as a bundle. When a prospect clicks a link for the newly recommended product, this action triggers a sequence of workflows in TIBCO Streaming software that re-estimate the lifetime value and generate a new quote.

The visual programming interface of TIBCO Streaming software exposes the business logic of a real-time event process. A new branch can be quickly implemented and integrated into the existing eventflow to cover new business use cases or extension scenarios.

Quote manipulation is a widespread and growing problem for insurers, brokers, and aggregators that can be easily handled by the TIBCO solution. Quote manipulation happens when customers deliberately omit or falsify information (such as making a female partner or elderly relative the primary driver, or constantly making modifications on parking locations, from on-street parking to a garage) to generate a lower premium. This impacts the cost base and revenue for insurers and pushes premiums up for genuine customers.

Figure 6. Quote manipulation analytics sub-branch.



TIBCO Streaming software can cache all incoming quote requests with the rules stored in its decision tables to identify the risk of driver fraud based on quoting behavior, such as constant change in parking locations from on-street parking to a secured car park, to private driveway, to garage; or quote risk attributes, for example, high risk area and vehicle with keyless entry technology. At point of quote, an automated workflow can be triggered to put potential fraudulent prospects into a black list, and their premium artificially inflated to encourage them to drop out. With a rich collection of adaptors available within the TIBCO Streaming system, and by using its ability to fuzzy match and link disparate customer records, you can connect and compare thousands of quotes obtained in a 90-day window from third-party broker engines or price comparison websites.

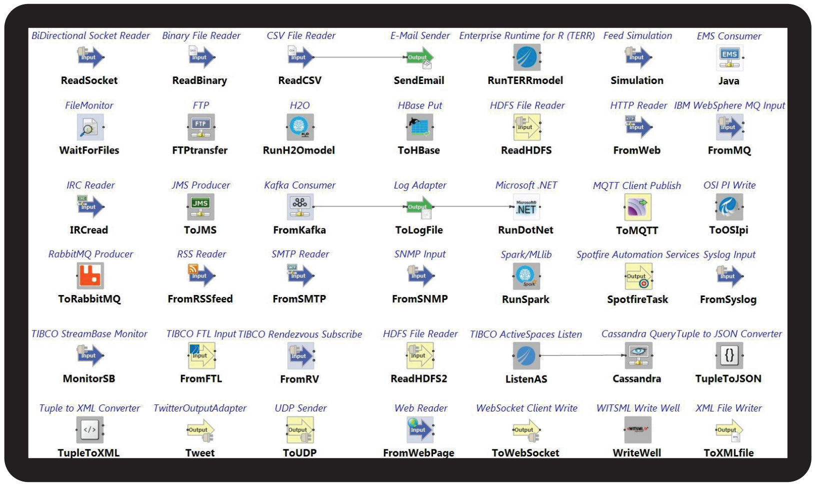
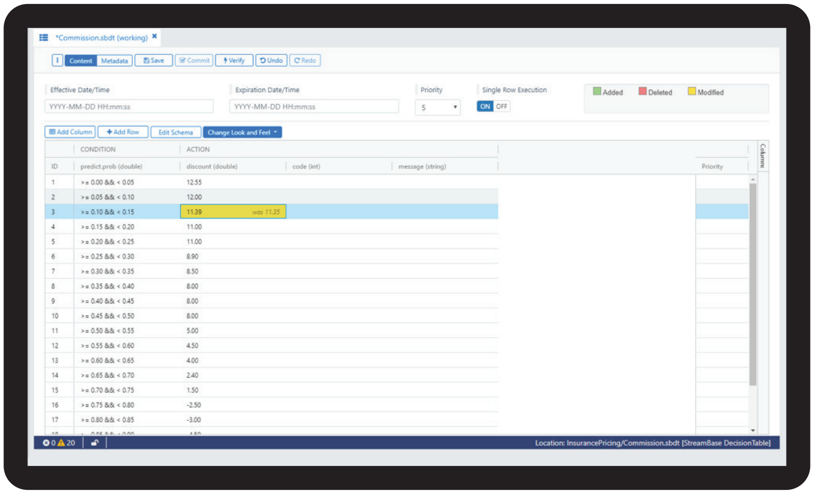


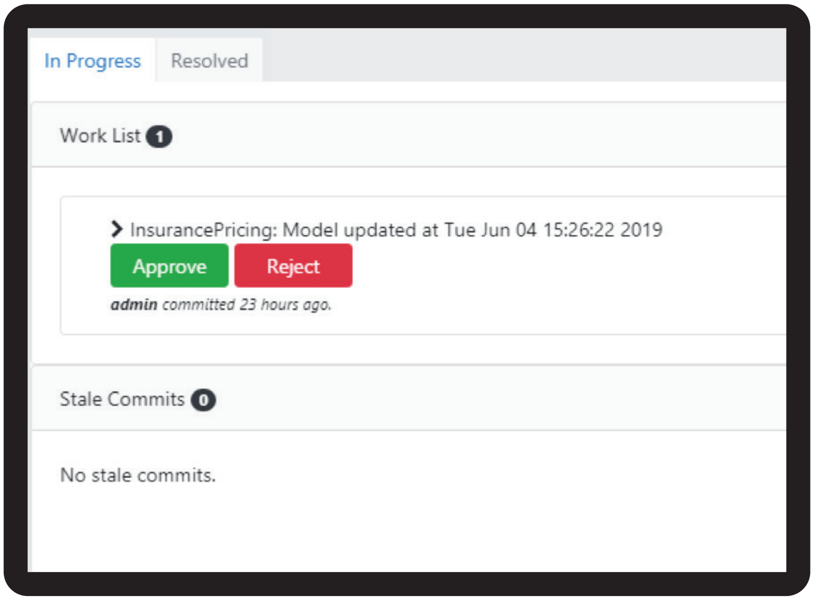
Figure 7. TIBCO Streaming pre-built adapters connect to real-time and historical data sources.

Competition plays a vital role in a real-time pricing application, and with TIBCO's solution, pricing analysts or actuaries can quickly dive into the web dashboard for detailed pricing analysis in the market and discover historical trends. Brand/category level indices can be monitored, and analysts can be notified using email/text alerts when certain patterns have been disrupted. This in turn can enable analysts to reset the competitive and profitable dynamic pricing rules within the TIBCO solution, as built using models from the competitive pricing intelligence engine.

For example, a sudden jump of conversion rate for a young driver group could represent the fact that the discount is too generous, and there may be a need to reduce the discount factor to protect the revenue streams. Alternatively, the latest request data can be fed back into the price elasticity model, allowing analysts to evaluate the supply/demand curve in near real time.

Figure 8. Sample of decision table that can be stored, controlled, and managed within the TIBCO Streaming system. A business user with the right privileges can log into the TIBCO Artifact Management Server and (a) quickly modify the discount factor, and (b) have the modification approved by their manager.





Summary

The TIBCO Connected Intelligence platform can be used to manage the entire algorithmic insurer insight-to-action cycle. The components of the solution include TIBCO Spotfire and TIBCO Data Science software for wrangling and analyzing historical quote data for price optimization, customer acquisition and retention; along with price elasticity data and modeling,

The models can be trained in memory, with GPU (CUDA or OpenCL) or in a big data lake, using TIBCO Data Science software. This includes inbuilt models, open source packages in R/Python as available in convenient notebooks and data functions, or in cloud-based ML libraries, such as Azure ML Services, AWS Sagemaker, and the Google AI platform.

After cross-validation, the champion model is deployed to the action components, which operate in real time on quote requests as they are submitted via TIBCO Streaming software. Each quote is scored, and a price and commission profile — fully customized for each individual customer — is obtained. The overall health of the system, as well as model effectiveness and explainability, can be evaluated in a real-time TIBCO Spotfire visual dashboard application. With the web-based interfaces involved, it is easy to enable registration, cross-validation, diagnostics, versioning, tracking, monitoring, and rebasing of analytical models to ensure they're performing well against the latest data.

Authors and Acknowledgements

This whitepaper was written by Michael O'Connell and Difei Luo, TIBCO Data Science Group. It refers to work done by other TIBCO staff.

The whitepaper benefits greatly from a number of TIBCO collaborations with the Automobile Association of Ireland. We acknowledge Colm Carey and Nina Jensen in this regard.

Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

20Sep2019