

Cofinanciado por:



UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

Designação do projeto | InsightZai – Machine Learning Modeling for Powerful Insights

Código do projeto | CENTRO-01-0247-FEDER-017728

Objetivo principal | Criação de uma plataforma completa de análise e modelação de grandes volumes de dados

Região de intervenção | CENTRO

Entidade beneficiária | Feedzai – Consultadoria e Inovação Tecnológica, S.A.

Data de aprovação | 2016-10-02

Data de início | 2016-07-01

Data de conclusão | 2018-06-30

Custo total elegível | 1.475.604,21€

Apoio financeiro da União Europeia | FEDER – 1.090.544,17€

Objetivos, atividades e resultados atingidos

O projeto "InsightZai" é implementado em co-promoção pela Feedzai e o Instituto Pedro Nunes (através do Laboratório de Informática e Sistemas, IPNIIS).

O objetivo fundamental do projeto "InsightZai" é criar uma plataforma completa de análise e modelação de grandes volumes de dados (*Data Science*), que a Feedzai aplicará, em primeira instância, à área da prevenção de fraude, mas que poderá ser usada, especialmente por clientes, em outros domínios. Simultaneamente esta plataforma deverá permitir uma transição unificada entre o mundo da modelação e o mundo da execução. Isto é, os modelos de prevenção de fraude criados devem ser possíveis executar diretamente em linha de autorização em sistemas de pagamentos.

De forma a concretizar os objetivos do projeto foram realizadas as seguintes atividades, do lado da Feedzai:

- Especificação de requisitos e arquitetura de alto nível;
- Implementação, configuração e validação de três versões (V1 - Modelação; V2 - Exportação de Modelos para Ambiente de Execução; V3 - Versão Final);
- Divulgação e promoção dos resultados.

O projeto "InsightZai" foi concluído com sucesso em Junho de 2018. Com as funcionalidades incluídas nos objetivos do projeto e alcançadas durante o desenvolvimento do projeto, a Feedzai passa, concretamente, a conseguir reduzir significativamente a velocidade de desenvolvimento de soluções de deteção de fraude e respetivo custo, destacando-se, com vantagem competitiva, dos seus principais concorrentes no mercado core da gestão de risco, em particular, na área financeira.



Acerca do projecto

O projecto "InsightZai" é implementado em co-promoção pela Feedzai e o Instituto Pedro Nunes (através do Laboratório de Informática e Sistemas, IPNlis).

Objetivo do projecto

O objetivo fundamental do projecto "InsightZai" é a criação de uma plataforma completa de análise e modelação de grandes volumes de dados que a Feedzai aplicará, em primeira instância, à área da prevenção de fraude, mas que poderá ser usada noutros domínios (e.g. na análise de risco na área dos seguros ou na verificação de fraude em pagamentos alternativos).

Simultaneamente, esta plataforma permite uma transição simplificada entre o mundo da modelação e treino de modelos (*Data Science*) e o mundo da execução (*runtime*). De forma unificada e completa, esta oferta não existe hoje no mercado.

Acerca da Feedzai

A Feedzai é uma empresa internacional que aplica técnicas de aprendizagem computacional (*machine learning*) e *Big Data* para prevenção de fraude bancária. Com foco nos ambientes de execução e um motor de aprendizagem e *scoring* em tempo real, a empresa é capaz de verificar mais de \$1500M em pagamentos diariamente, o que corresponde a mais de 20M transacções/dia, em tempo real.

INSIGHTZAI

PLATAFORMA COMPLETA DE ANÁLISE E MODELAÇÃO DE GRANDES VOLUMES DE DADOS

O "InsightZai" é um protótipo de uma solução completa e integrada para modelação, resultado da vasta experiência da Feedzai na área da ciência de dados e aprendizagem computacional, complementada com a colaboração do IPNlis ao nível do estudo e da criação de soluções para a melhoria da experiência de interacção do utilizador.

Objetivos do projecto

- 1 Garantir que o "InsightZai" é uma solução completa e integrada para modelação e prevenção de fraude usando técnicas de aprendizagem computacional e *Big Data*.
- 2 Encurtar a duração das iterações da fase de exploração de dados, aumentando a eficiência e eficácia dos cientistas de dados. Através desta plataforma, processos que atualmente demoram vários meses deverão ser reduzidos a semanas.
- 3 Relaxar o conjunto de competências exigidas aos cientistas de dados, democratizando o acesso aos benefícios da ciência de dados e aprendizagem computacional a novos clientes e mercados.
- 4 Realizar uma validação experimental dos resultados obtidos em ambiente controlado e ambiente real, indo de encontro ao interesse de alguns clientes na solução.
- 5 Aumentar a capacidade exportadora da Feedzai para o mercado mundial, criando uma oferta competitiva e única na aplicação da ciência de dados à deteção de fraude.
- 6 Melhorar a competitividade da indústria portuguesa na área das Tecnologias da Informação e, em particular, na área de Deteção de Fraude com recurso a aprendizagem computacional, gerando conhecimento e valor que pode ser capitalizado e exportado para o mercado mundial.

Cofinanciado por:



Fundos Europeus Estruturais e de Investimento

Desenvolvimento do projecto

A construção dos **módulos do projecto** resultou de uma criteriosa fase de levantamento de requisitos com o envolvimento de utilizadores finais do sistema, seguida pela fase de definição da arquitetura do sistema usando uma abordagem ágil de experimentação e prototipagem rápida.

Caracterização de fontes de dados

1 Datasources

Este módulo permite o carregamento e/ou o **acesso a grandes quantidades de dados** para que possam ser usados corretamente e eficientemente pelo sistema.

Algumas das funcionalidades permitidas são: cálculo e visualização de estatísticas sobre fontes de dados e também a integração com plataforma Hadoop/HDFS.

Plano de dados para amostragem estratificada

2 Data Science Plans

Com este módulo, a fase de **transformação, limpeza e enriquecimento das fontes de dados** (geralmente chamada de *feature engineering*) torna-se mais simples, graças ao interface mais *user friendly*.

De igual forma, a integração com plataformas de processamento distribuído (Hadoop e Spark) torna-se mais transparente.

Parametrização do algoritmo Random Forest (proprietário da Feedzai)

3 Model Projects

A **integração de modelos de machine learning** e a possibilidade de parametrização dos mesmos ao nível da plataforma é realizada neste módulo, bem como o processo de treino do algoritmo com as fontes de dados.

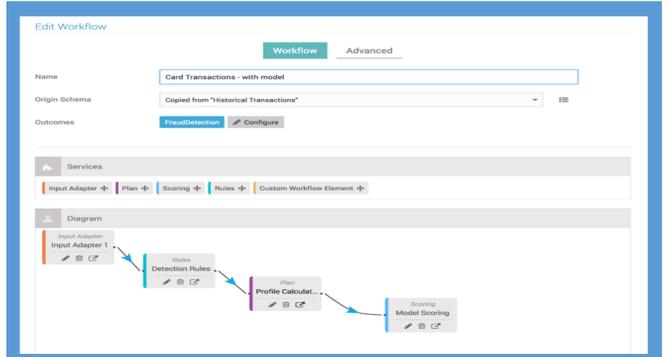
Os resultados dos testes de modelos poderão ser consultados num relatório de desempenho, que demonstra as principais métricas de desempenho e de precisão do modelo.

Cofinanciado por:

4 Workflows

A passagem de um modelo de *data science* para *runtime* requer a criação de um *workflow* onde se configura a seqüência através da qual o sistema Pulse deverá tratar os dados reais recebidos de uma fonte externa.

Este módulo permite a **total integração do ambiente de exploração de dados e do ambiente de execução de modelos**, bem como as suas dependências.



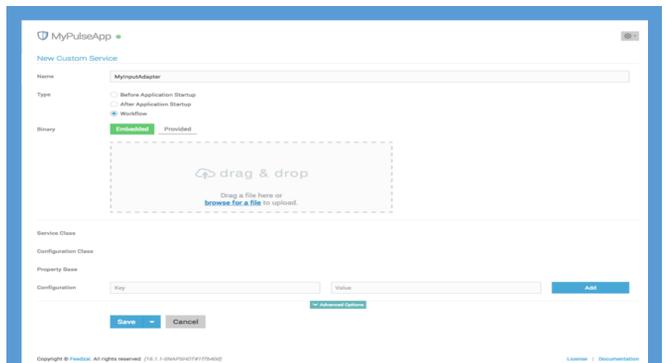
The screenshot shows the 'Edit Workflow' interface. At the top, there are tabs for 'Workflow' and 'Advanced'. The 'Name' field is 'Card Transactions - with model'. Below, there are sections for 'Services' and 'Diagram'. The 'Diagram' section shows a flowchart with four main components: 'Input Adapter', 'Detection Rules', 'Profile Calculat...', and 'Scoring Model Scoring', connected by arrows indicating the flow of data.

Configuração do Workflow

5 Custom Services

O *Workflow* permite definir o **processamento de um evento em ambiente de execução** através de uma *interface* que permite construir um grafo de serviços.

Para que estes serviços sejam compatíveis com o *Workflow* é necessário implementar uma API pública fornecida pela Feedzai aos seus clientes. Este módulo permite a **importação dos serviços** na plataforma.



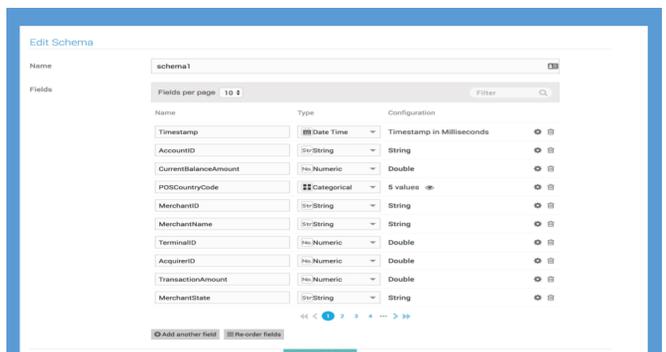
The screenshot shows the 'New Custom Service' interface. It includes fields for 'Name' (MyInputAdapter), 'Type' (Workflow), and 'Binary' (Embedded). There is a 'drag & drop' area with instructions. Below, there are sections for 'Service Class', 'Configuration Class', 'Property Base', and 'Configuration'. A 'Save' button is visible at the bottom.

Configuração de serviços customizados do Workflow

6 Schemas

Um *schema* representa um conjunto de campos e a sua configuração, permitindo aos conceitos que o usam definirem o contrato pelo qual “comunicam” entre si.

No *workflow*, tem de existir pelo menos um **serviço cujo objectivo é produzir o evento em si** que será processado pelos restantes serviços a jusante. O *schema* destes serviços são criados nesta interface própria.



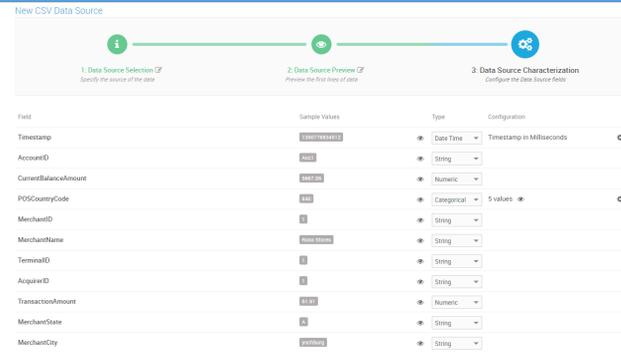
The screenshot shows the 'Edit Schema' interface. It displays a table of fields for a schema named 'schema1'. The table has columns for 'Name', 'Type', and 'Configuration'. The fields listed are: Timestamp (Date Time), AccountID (String), CurrentBalanceAmount (Numeric), POSCountryCode (Categorical), MerchantID (String), MerchantName (String), TerminalID (Numeric), AcquirerID (Numeric), TransactionAmount (Numeric), and MerchantState (String).

Configuração de Schemas

Cofinanciado por:



FUNDO EUROPEO DE INVESTIMENTO

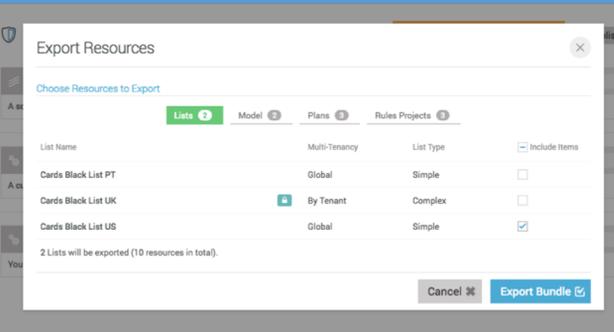


Definição de métricas e indicadores finais

7 Sandbox evaluation

Este módulo permite **avaliar modelos e regras em paralelo antes da sua migração para ambiente de produção**, tal como acontece em ambientes reais.

As *sandbox evaluations* permitem a criação de um ambiente semelhante a um *workflow*, possibilitando a simulação de um ambiente composto por regras e modelos que **contribuem em simultâneo para um decisão final** por transação ou evento.



Seleção e configuração de recursos a exportar

8 App Collaboration

Neste módulo, o **trabalho realizado em dado ambiente pode ser exportado**, permitindo ao cientista de dados **importar para outro ambiente**, desde que tenha acesso a esse ambiente de destino.

Esta funcionalidade permite que diversos utilizadores **colaborem entre vários ambientes de desenvolvimento**. Em caso de conflitos entre o ambiente de partida e destino, é possível resolver esses conflitos num dos passos de importação.

Conclusão

O projecto "InsightZai" permitirá à Feedzai aumentar o número de funcionalidades do seu principal produto "Pulse", permitindo-lhe expandir a sua atividade a nível mundial. Com as funcionalidades que estão incluídas no objectivo do projecto, a Feedzai irá conseguir reduzir significativamente o custo e velocidade de desenvolvimento de soluções de detecção de fraude, factor essencial para reagir aos desafios atuais desta área.

Cofinanciado por: