

Relatório do trabalho da disciplina de Integração de Sistemas de
Informação

ETL de dados de carros de *drift* com KNIME

Carlos Filipe Santos Brandão - 26025

LESI-PL

Outubro de 202

Índice

1.	ENQUADRAMENTO	3
2.	PROBLEMA	3
3.	ESTRATÉGIA UTILIZADA	4
4.	TRANSFORMAÇÕES	5
5.	VIDEO COM DEMONSTRAÇÃO	5
6.	CONCLUSÃO	6
7.	REFERÊNCIAS BIBLIOGRÁFICAS E WEB	6

1. Enquadramento

Equipa: Trabalho Individual de Carlos Filipe Santos Brandão

Unidade Curricular: Integração de Sistemas de informação (ISI)

Curso: LESI-PL

Breve descrição: o projeto utiliza exclusivamente a plataforma KNIME para implementar um processo de ETL que lê um ficheiro .xlsx com dados sobre carros de drift (Modelo, Potência, Largura dos Pneus, Estado, etc.), efetua transformações, gera ficheiros de saída em vários formatos, organiza-os em pastas, comprime num zip e envia por email.

2. Problema

O que se pretende demonstrar / resolver:

- Demonstrar a construção de um pipeline ETL completo usando KNIME para processamento de dados de carros de drift;
- Limpar, normalizar e transformar os dados para produzir relatórios e ficheiros JSON prontos a integrar noutros sistemas;
- Organização programática dos ficheiros de saída em pastas por critério (ex.: separação de veículos por classe/grupo, aprovação, veículos eliminados, *score* de relação de potência com aderência, etc);
- Automação do empacotamento (zip) e entrega por email;
- Garantir regras de negócio aplicadas (ex.: categorização segundo potência, validação de campos).

Motivação prática: replicar um cenário real onde um departamento técnico/operacional recolhe dados de veículos e necessita de um pipeline repetível para preparar dados para análises, partilha e integração com sistemas externos.

3. Estratégia utilizada

Ferramenta escolhida: KNIME Analytics Platform (workflow visual de nós)

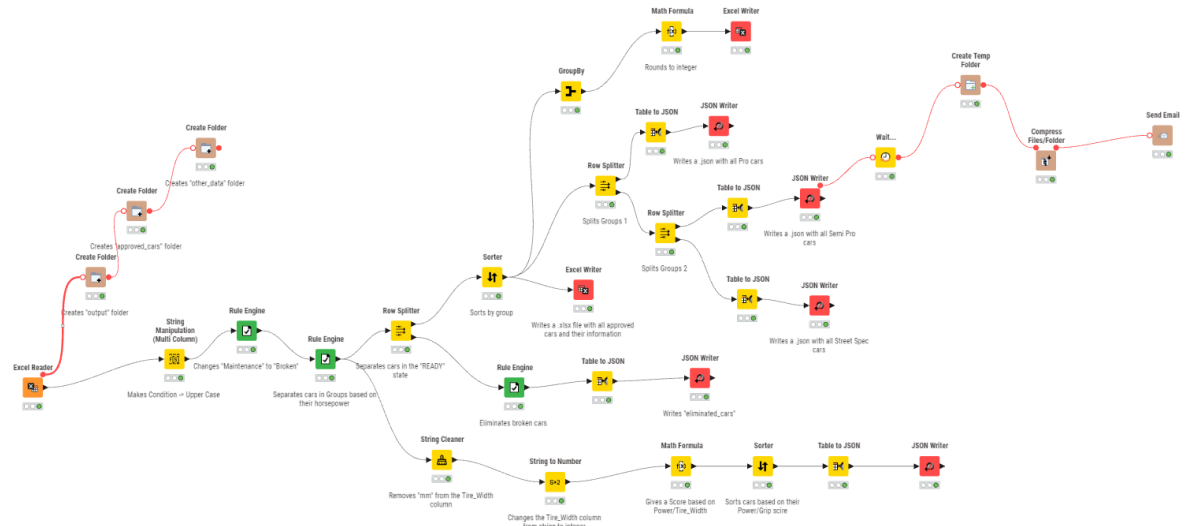
Fluxo geral (alto nível):

4. Leitura do ficheiro .xlsx (Excel Reader);
5. Pré-validação e limpeza (String Cleaner, Row Filter, Rule Engine);
6. Ordenação e segmentação (Sorter, Row Splitter);
7. Cálculos e normalizações (Math Formula);
8. Agrupamentos e agregações (GroupBy);
9. Serialização/Export (Table to JSON);
10. Organização de ficheiros (Create Folder);
11. Compressão (Zip Files);
12. Envio por email (Send Email).

Operadores / Nodes KNIME usados:

1. Excel Reader;
2. String Manipulation / String Cleaner;
3. Row Filter;
4. Sorter;
5. Row Splitter;
6. Math Formula;
7. Rule Engine;
8. GroupBy;
9. Table to JSON;
10. JSON Writer;
11. Create Folder / Zip Files (nodes de arquivos);
12. Send Email.

4. Transformações



Explicação das transformações passo-a-passo:

5. Video com demonstração

https://drive.google.com/file/d/1bg8V9_Gfi7Q63Yp2i3z5SZfFc_v8oEs4/view?usp=sharing

6. Conclusão

O workflow em KNIME demonstrou a viabilidade de construir um pipeline ETL robusto e repetível para dados de carros de drift. O processo cobre a leitura de ficheiros Excel, limpeza, validação, segmentação, cálculo de métricas, escrita em vários formatos, organização de ficheiros, compressão e envio por email, mantendo registos (logs) para auditoria.

Trabalhos futuros / melhorias que queria fazer:

- Integração com APIs externas;
- Dashboard interativo (ex.: conectar os resultados a um painel em Grafana/Power BI/Tableau ou usar KNIME WebPortal);
- Testes unitários para as regras de transformação e pipelines (ex.: validar outputs esperados para ficheiros de input de teste);
- Operações com bases de dados SQL.

7. Referências bibliográficas e web

- Plataforma de e-learning Moodle (materiais da UC e exemplos de projetos anteriores);
- Tutoriais do YouTube sobre criação de workflows e automações no KNIME;
- Documentação oficial do KNIME (<https://docs.knime.com>) — consulta de detalhes técnicos sobre os nodes usados e boas práticas de ETL.