

Universidade do Minho

Aplicações Informáticas na Biomedicina
Ficha 5 - Grupo 3

Working Group: Filipe Fortunato - A75008
João Leal - A75569
José Sousa - A74678
Ricardo Canela - A74568

19 de Novembro de 2019

Conteúdo

1	Introdução	3
2	Jobs	3
2.1	Job 1	3
2.2	Job 2	4
2.3	Job 3	4
3	Conclusão	6

1 Introdução

Esta ficha foi realizada no âmbito da disciplina de *Aplicações Informáticas na Biomedicina* e foi-nos pedido que desenvolvêssemos 3 Jobs usando a ferramenta *Talend*. Para a criação destes Jobs os docentes forneceram-nos um dataset **mental.health.csv**. Este dataset contém dados de um estudo realizado em 2014 em são avaliadas algumas atitudes dos trabalhadores em empresas de Tecnologias de Informação (TIs) relativamente à saúde mental, bem como a frequência de transtornos mentais em ambientes de trabalho.

2 Jobs

Nesta secção serão apresentados os 3 Jobs criados por nós. Iremos explicar o porquê da criação da Job e os diferentes passos para realização dos mesmos.

2.1 Job 1

Neste Job pretendemos determinar quais são os trabalhadores que já receberam tratamento para doenças mentais e que têm histórico familiar de doenças mentais, separados por género.

Para isto resolvermos guardar os seguintes dados dos utentes:

- Age
- Gender
- Country
- State
- self_employed
- family_history
- treatment

Em primeiro lugar, criamos a ligação à base de dados e também importamos o dataset. De seguida filtramos quais eram os utentes que já receberam tratamento e têm histórico familiar de doenças mentais usando as variáveis *treatment* e *family_history*. De seguida, separamos esses utentes por género usando a variável *Gender*. Para finalizar, criamos uma tabela para cada género. Optamos também por criar uma tabela com os trabalhadores que não receberam tratamento.

Sendo assim, após o tratamento dos dados ficamos com as seguintes tabelas :

- female_tratamento
- male_Tratamento
- sem_tratamento

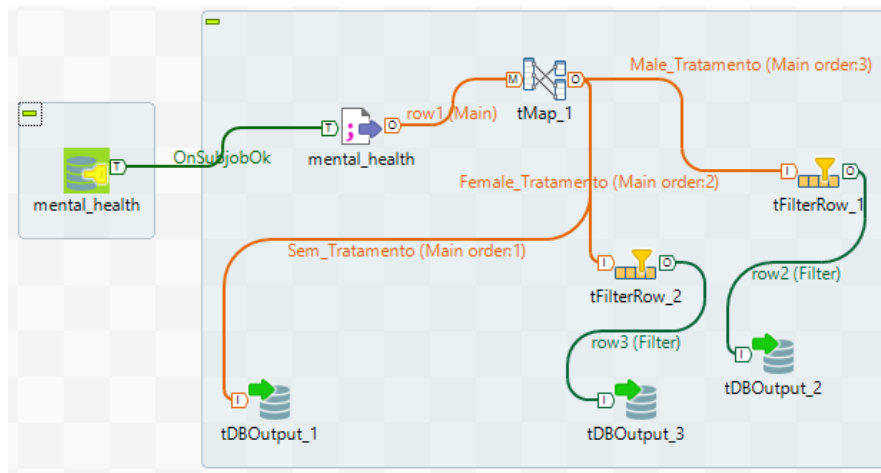


Figura 1: Esquema *job1*.

2.2 Job 2

Neste *Job* pretendemos determinar quais as pessoas que têm empresa própria e já receberam tratamento para doenças mentais, ordenando-os pela idade. Foram guardados todos os dados dos diferentes utentes.

Em primeiro lugar, criamos a ligação à base de dados e também importamos o dataset. De seguida começamos por ordenar os trabalhadores pelas suas idades. De seguida aplicamos um filtro para determinar quais são os que têm empresa própria. Por fim guardamos todos os dados desses trabalhadores num ficheiro Excel.

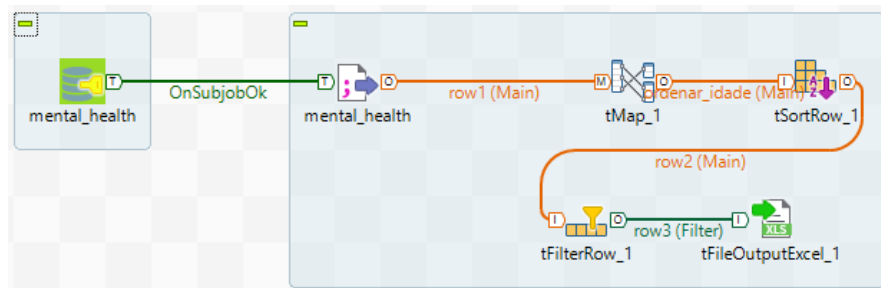


Figura 2: Esquema *job2*.

2.3 Job 3

Este *job* permite identificar os trabalhadores com facilidade, ou não, de falar sobre doenças mentais com o seus colegas de trabalho ou supervisores e em ambas as situações determinar aqueles que procuraram tratamento, ou preferem não o fazer.

Para a construção do *Job* precisamos, inicialmente, de efetuar a conexão com a base de dados e o consequente carregamento de dados. Nesta fase é efetuado o tratamento de dados e este é feito em dois momentos. Os dados são separados de

acordo com os valores das variáveis *coworker* e *supervisor*, ou seja, distinguir os indivíduos que têm, ou não, facilidades em conversar sobre doenças mentais. Os indivíduos separados, são novamente filtrados de acordo com a sua opção de já terem, ou não, procurado algum tipo de tratamento.

São então obtidas, do tratamento de dados, quatro tabelas diferentes:

- Partilhar & Tratamento
- Partilhar & Não Tratamento
- Não Partilhar & Tratamento
- Não Partilhar & Não Tratamento

As informações dos trabalhadores, foram guardadas de forma imediata, há medida que eram necessárias no tratamento dos dados.

- *Treatment*
- *Coworker*
- *Supervisor*

Para além destas, resolvemos guardar outras variáveis de forma a facilitar uma análise futura.

- *Age*
- *Care Options*
- *Wellness Program*
- *Anonymity*
- *Mental health consequence*
- *Mental health interview*

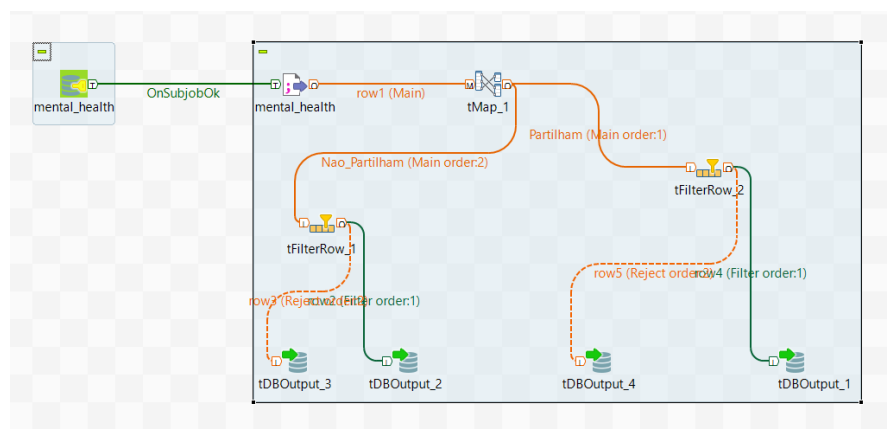


Figura 3: Esquema *job3*.

3 Conclusão

Ao concluir a resolução dos exemplos propostos conseguimos adquirir uma melhor percepção dos métodos de processamento de um conjunto de dados e, de forma complementar, percepção sobre os diferentes resultados possíveis de obter.

A ferramenta, *Talend*, foi importante na resolução dos exemplos permitindo de uma forma, relativamente, simples a criação e ligação dos diferentes componentes úteis ao tratamento de dados (ex: *tMap*, *tFilterRow*, etc.)

Foram portanto adquiridas boas práticas de processamento de um *dataset* assim como experiência com uma ferramenta fundamental para tal, como o *Talend*.