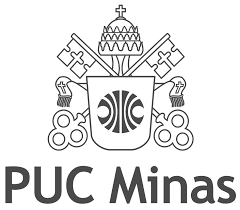
****

**PÓS GRADUÇÃO EM CIÊNCIA DE DADOS E BIG DATA**

**Disciplina:** Aprendizado de Máquina

**Professor:** Sérgio M. Dias

**Alunos:** Alessandra Araújo de Oliveira, Rodrigo Pimentel dos Santos.

**RELATÓRIO TÉCNICO**

**- Criação de Modelo de recomendação para e-commerce de calçados -**

1. **Resumo da proposta**

Entender o perfil de consumo dos clientes do e-commerce da marca Luiza Barcelos e analisar a possibilidade de criação de um modelo de recomendação de produtos para clientes já existes na base de dados da empresa. A construção desse modelo partiria da análise das compras realizadas no próprio sita da marca por clientes com características similares.

1. **Introdução**

A Luiza Barcelos é uma marca de calçados de moda com trinta anos de mercado que atua em diversos canais no Varejo. Seus produtos são comercializados para lojistas multimarcas, varejo próprio - por meio de suas lojas e franquias - e através de seu e-commerce. A marca vem passando por um período de crescimento expressivo, chegando dobrar seu faturamento no ano de 2019.

O e-commerce da marca é responsável por aproximadamente 35% das vendas do canal próprio no *Sell Out* e apresenta grande projeção de crescimento devido à mudança de comportamento dos consumidores com relação à compra de produtos online e à esperada melhora do cenário econômico do país.

Dentro desse contexto, existe grande interesse da empresa em entender melhor os padrões de comportamento dos clientes online e utilizar a tecnologia para elaborar formas de aumentar sua taxa de conversão neste canal e explorar ao máximo seu potencial de crescimento.

1. **Motivação**

A necessidade de entender os diversos perfis de clientes que compõe o e-commerce da marca e avaliar a viabilidade de criação um sistema de recomendação de produtos para os clientes de acordo com o consumo de clientes com características e preferências similares.

1. **Perguntas**

Existe correlação entre o perfil de consumo dos clientes da marca? Essa correlação existente é alta ou baixa?

1. **Hipóteses**

Dentre as perguntas criadas acima, podemos criar as seguintes hipóteses:

Não existe nenhum padrão de relacionamento entre a compra dos clientes do e-commerce Luiza Barcelos, não possibilitando criação de um modelo eficiente que faça sugestão de produtos com base nessa relação.

Existe grande padrão de consumo das clientes que fazem compra no site da marca, o que possibilita a criação de um modelo eficiente que faça sugestão de produtos com base nessa relação.

Existe moderada relação entre o padrão de consumo das clientes que fazem compra no site da marca, o que possibilita a criação de um modelo que faça sugestão de produtos com base nessa relação, porém com baixa confiabilidade.

Existe relação entre o padrão de consumo para determinados produtos, o que possibilita a criação de modelos de recomendação apenas a partir da compra desses itens.

1. **Metodologia**

Para criar recomendações de produtos através do histórico de compra dos clientes da empresa, iremos utilizar aplicação de algoritmos para fazer descobertas de regras de associação. A aplicação de regras de associação consiste no processo de analisar os relacionamentos existentes entre atributos de base de dados transacional, com o objetivo de encontrar associações ou correlações.

Os algoritmos utilizados para descobrir as regras de associação se baseiam no pressuposto de que a presença de um atributo X no evento implica na presença de atributo Y em um mesmo evento. As regras de associação geralmente são representadas por meio de afirmações: *SE*.... *ENTÃO*.....

Para se descobrir a regra de associação de dados transacionais, independente do algoritmo utilizado, deve-se criar um conjunto com os itens de cada transação da base de dados. Com essa transformação, teremos vários conjuntos, cada um deles composto pelos itens de determinada transação. Um conjunto de itens, ou um subconjunto de itens do domínio, é denominado itemset.

Para validar as regras de associação apresentadas pelos algoritmos são utilizadas as medidas de suporte e confiança. A medida de suporte, quando aplicada a um itemset diz respeito à frequência desse item na base transacional sob análise. Já a medida de confiança é aplicada somente às regras criadas pelos modelos de associação e ela indica a frequência em que a regra ocorre mediante premissa definida.

Além dos indicadores citados acima, utiliza-se para melhorar a análise das regras de associação a medida de correlação Lift, que indica o grau de correlação de cada regra, sendo:

* Lift > 1 – correlação positiva
* Lift = 1 – descorrelação
* Lift < 1 – correlação negativa

1. **Experimentos**

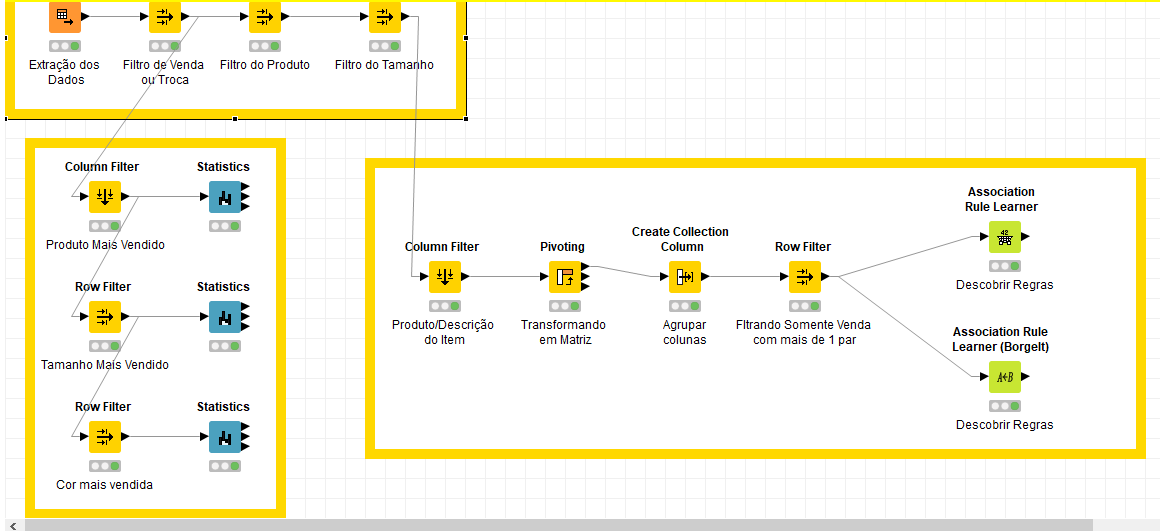
Para entender se existe algum padrão de compra dos clientes do e-commerce da marca Luiza Barcelos foi extraída uma base de dados com as vendas do site em um período de 6 meses (de julho à dezembro de 2019). A base de dados extraída continha as seguintes variáveis:

* Código do cliente
* Nome do cliente
* UF
* Filial
* Data faturamento
* Código do produto
* Descrição do produto
* Tamanho
* Cor

Os primeiros experimentos com os dados falharam, uma vez que tínhamos um número muito grande de quantidade de produtos as medidas de suporte ficaram com valores extremamente baixos. Após os as primeiras análises identificamos a necessidade de segmentar melhor os dados para conseguir identificar melhor os resultados.

Após uma análise descritiva dos dados, foi possível propor dois modelos de recomendação de produtos. Os dois modelos propostos são resultado da análise descritiva de frequência dos produtos mais vendidos, tamanho, cores e tipos.

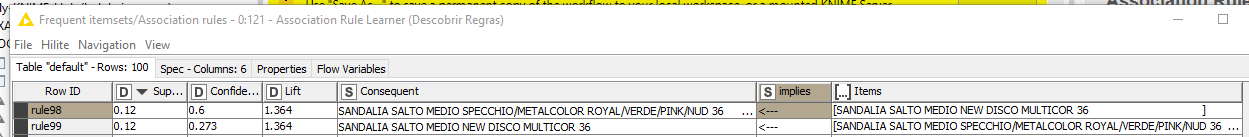
No primeiro fluxo, a base de dados foi carregada e, na sequência foi filtrado apenas um único produto - o produto com a melhor posição do ranking de vendas da marca - para identificar a possibilidade de recomendar compra de uma segunda cor desse produto durante a compra de uma cor de tal item, tendo em vista a grande popularidade deste produto.

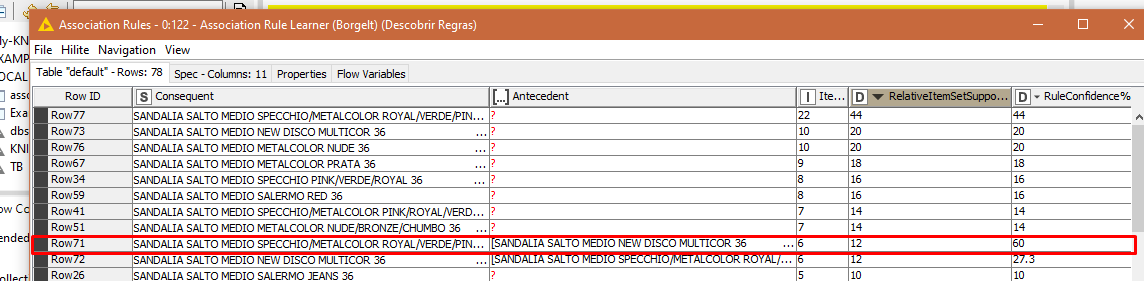


Fluxo de modelo de associação software Knime \_ v1

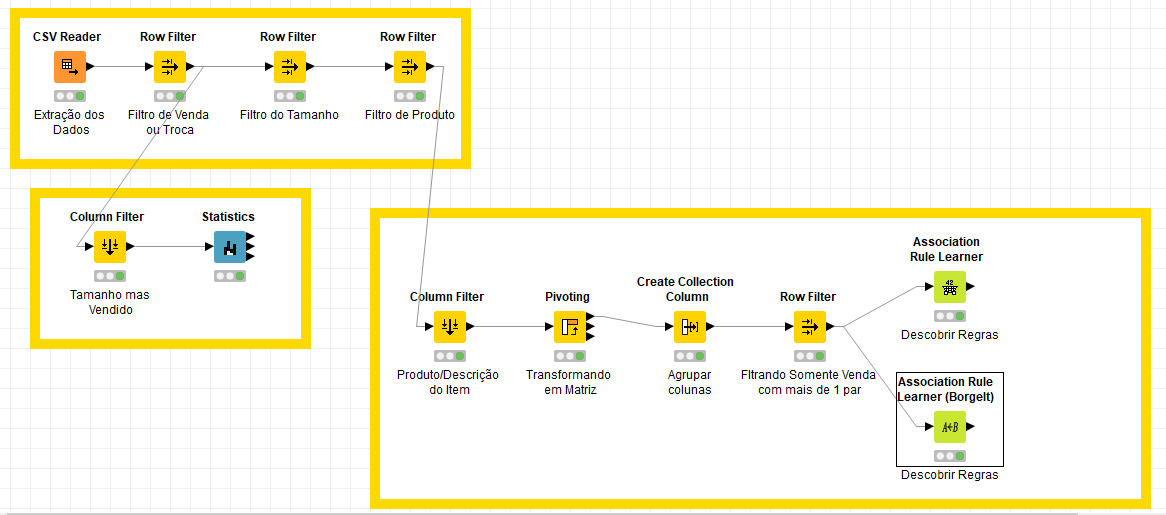
Após rodar o modelo, foi observado que a cor Specchio Metalcolor Royal foi comprada em 12% das ocorrências (medida de suporte), e que, nesses casos, a cor New Disco multicor também era comprada em 60% dos casos de compra do Specchio Metalcolor Royal (medida de confiança). O Lift apresentado nessa regra foi 1,364, o que indica uma correlação positiva entre essas duas cores desse produto.

Segue abaixo a regra apresentada pelo software knime após rodar os modelos de associação.





No segundo fluxo, a base de dados foi carregada e, na sequência foram filtradas somente a compra de calçados tipo Bota (uma única categoria) com o intuito de identificar qual produto deve ser recomendado para clientes na compra de determinado item.



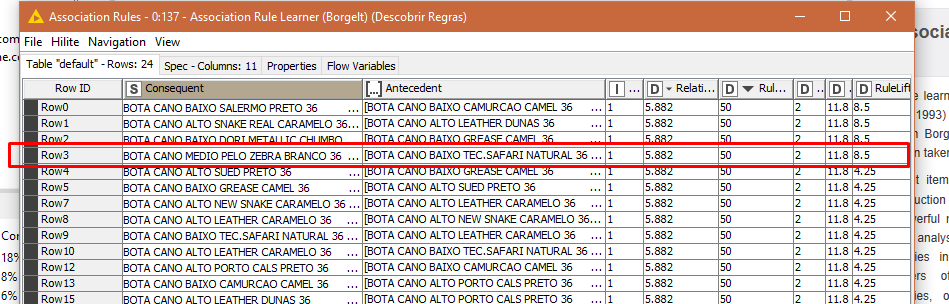
Fluxo de modelo de associação software Knime \_ v2

Após rodar o modelo, foi observado que em 5,9% das ocorrências de compra de mais de um produto por um mesmo cliente (medida de suporte), quando um cliente comprava a “Bota Cano Baixo Tec. Safari Natural” ele também comprava a “Bota Cano médio pelo Zebra Branco” 100% dos casos (medida de confiança). O Lift apresentado nessa regra foi 8,5, o que indica uma altíssima correlação positiva entre esses dois produtos.

Ao tratar os dados nesse segundo exemplo, a base de dados ficou muito limitada, com poucas linhas, o que nos faz concluir que a amostra utilizada foi muito pequena para realizar uma análise que, de fato, represente e simule a realidade.

Segue abaixo a regra apresentada pelo software knime após rodar os modelos de associação.





1. **Conclusões**

Após aplicação dos modelos de associação na base de dados transacionais do e-commerce da Luiza Barcelos, conclui-se que é possível extrair regras coerentes o que diz respeito ao consumo dos produtos da marca. No entanto, para conseguir efetuar a extração de regras que, de fato, fazem sentido, é preciso efetuar tratativas na base como, por exemplo, segmentar a base de dados por categoria de produto. Como são muitos produtos diferentes comprados e fabricados pela empresa, é complicado encontrar regras de associação com alto valor de suporte com uma base de dados é muito pulverizada no que diz respeito à variedade dos produtos.

Para encontrar regras que fizessem mais sentido e apresentassem um valor de suporte mais consistente, tivemos que filtrar a base de dados, considerando somente um modelo de sapato (referência) para identificar qual cor desse produto sugerir em detrimento de compra prévia determinada cor desse mesmo item. Tal modelagem também poderia ser efetuada agrupando os produtos em categorias mais genéricas (considerando a modelagem do calçado por exemplo) para realizar a aplicação das regras de associação considerando a categoria de produto e não o código do produto em sim, dessa forma conseguira-se extrair regras semelhantes ao exemplo abaixo:

Clientes que compram sapatilha geralmente também irão comprar sandália rasteira.

Regras como a descrita acima, mesmo que mais genéricas, possibilitariam mais opções de sugestões de produtos para os clientes.

Os próximos passos para esse projeto são a melhoria da segmentação para criação de novos fluxos, a aplicação da técnica no e-commerce e aplicação da técnica para sugestão de vendas dos vendedores e representantes da empresa.