

Histórico do assunto

A Varejo Company é uma empresa fictícia, porém, baseada em uma empresa real. Ela está atuando no mercado de alimentação há mais de 50 anos mas atualmente tem enfrentado os mais variados tipos de problemas por conta da sua estrutura defasada.

Isto ocorre porque apesar de hoje ser líder em seu segmento e ter um faturamento de mais de 100 milhões anuais, ainda mantêm muito de seus processos da mesma forma que foi desenhado no início da companhia e, por isto, tem enfrentado diversos pontos de insatisfação dos seus clientes como queda na qualidade de seus produtos, atraso nas entregas, marketing pouco efetivo, baixa personalização para os mercados regionais, entre outros que estão acarretando na perda de mercado para novos concorrentes.

A empresa conta atualmente com 112 vendedores sendo:

- 17 vendedores de atendimento externo (presencial) de Nível 3 com aproximadamente 100 clientes na carteira de cada um deles;
- 37 vendedores de atendimento externo (presencial) de Nível 2 com aproximadamente 200 clientes na carteira de cada um deles;
- 58 vendedores de atendimento interno (email/telefone) de Nível 1 com aproximadamente 500 clientes na carteira de cada um deles;

Apesar de ser uma empresa criada, os problemas listados e a base de dados trabalhada são condizentes à desafios reais que tive que resolver em uma época da minha vida corporativa.

Descrição do problema

Para reverter esta situação a empresa contratou diversas equipes de consultores onde cada uma terá um problema a ser resolvido. Nós da Udacity Team somos responsáveis por garantir uma melhor eficiência nas entregas, diminuindo os custos de transporte e aumentando a satisfação dos clientes. Ao entrevistar a equipe logística já foi possível entender alguns problemas. Os principais são:

1. Apesar do porte da empresa influenciar no tratamento oferecido pela Varejo Company, não há regras para classificá-la em pequeno, médio ou grande porte. Ficando a cargo do gestor da região interpretar e escolher qual segmento ele considera mais apropriado;
2. Como a classificação de segmentação está bagunçada, a equipe logística não tem uma periodicidade de entregas personalizada para cada porte de empresa, ocasionando frequências alta de entrega para clientes pequenos e atrasos nos clientes grandes;
3. A equipe se torna mal dimensionada uma vez que os Vendedores de nível 3 (maior experiência e consequentemente salários altos) acabam perdendo muito tempo com clientes pequenos que poderiam estar sendo atendidos por vendedores de nível 1 (menor experiência e salário base) acarretando em uma folha salarial inflacionada.

Conjuntos de dados e entradas

Para auxiliar na análise a empresa nos forneceu uma base de dados em Excel com o histórico de compra de seus 37.765 clientes. Nesta planilha temos disponíveis as seguintes Features:

- **Gestor:** nome do gestor responsável pelo cliente;
- **Regiões de Venda:** região que o cliente se encontra entre as 4 regiões definidas pela empresa;
- **CD:** centro de distribuição responsável pela entrega;

- **Porte:** porte da empresa que foi definido pelo vendedor (grande, médio ou pequeno);
- **Vendedor:** vendedor responsável pelo cliente;
- **Cliente:** nome do cliente;
- **Vol TT:** volume total dos últimos 2 anos em tonelada;
- **Vol Méd:** média de compra dos últimos 2 anos em tonelada;
- **Freq Compra:** dado medido em percentual que nos diz quantos meses o cliente realizou uma compra em relação aos 24 meses analisados. Por exemplo, se o cliente tem uma frequência de 50%, quer dizer que ele realizou alguma movimentação em 12 meses do período.

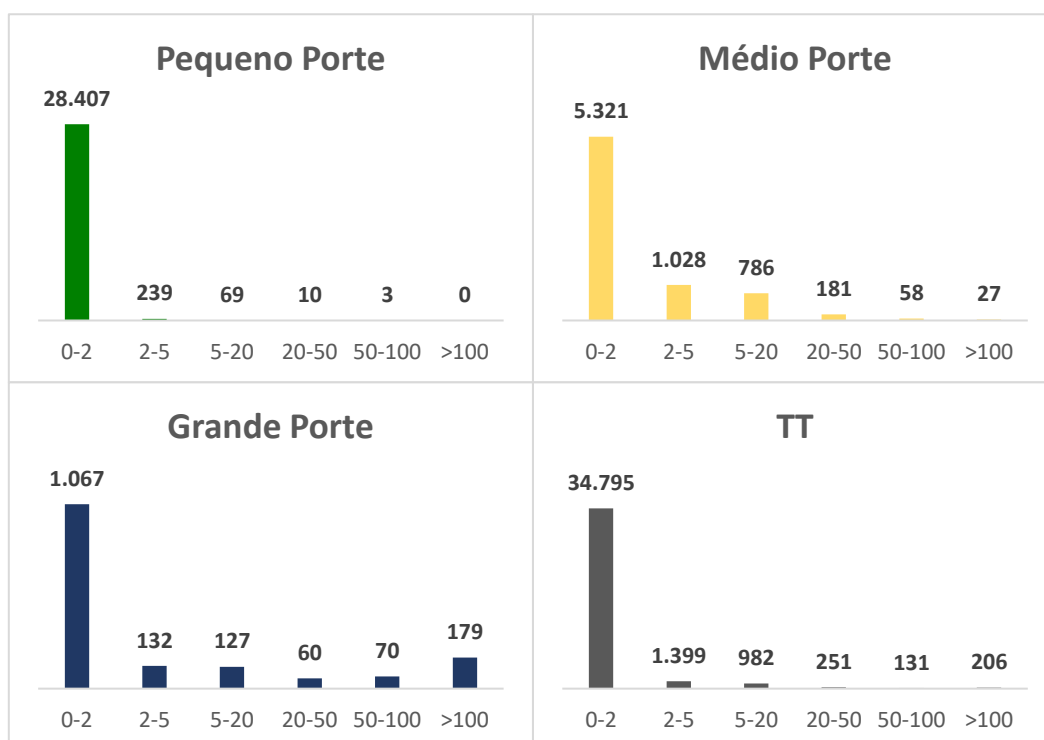
Descrição da solução

Para resolver os problemas listados primeiramente será feito um modelo de aprendizagem não supervisionada para definir as linhas de corte para a clusterização dos 3 segmentos (grande, médio, pequeno). Após isto, um modelo de aprendizagem supervisionada será treinado para classificar os clientes no seu segmento correto.

Este modelo será integrado com a base de clientes para que a cada novo cliente conquistado, ele já seja direcionado diretamente para o segmento correto, retirando este poder de decisão da mão do vendedor. Para garantir a classificação correta dos clientes nós utilizaremos como métrica o f-score.

Modelo de referência (benchmark)

Como podemos ver no gráfico abaixo não há um padrão de volume para definir o porte da empresa, ocasionando em clientes dos mais diversos portes em todos os segmentos. Por exemplo, no segmento de Médio Porte temos 5.231 clientes que compram numa faixa de 2 a 5 toneladas mensais ao mesmo tempo que abriga também 27 clientes que compra mais de 100 tons mensalmente.



Métricas de avaliação

Para definir qual será o volume mínimo (em tonelada) para pertencer a cada segmento adotaremos o conceito de Curva ABC.

Este método irá auxiliar na separação dos clientes de maior impacto para a empresa dos menores da seguinte forma: os clientes serão classificados em ordem crescente de volume, os primeiros clientes (os menores) cujo a soma de seus respectivos volumes totalizem aproximadamente 80% serão classificados como Pequeno Porte, os próximos 15% como Médio Porte e os últimos 5% como Grande Porte.

$$A = \frac{\sum iVol \text{ Médio Clientes Grande Porte}}{Volume Total} \cong 80\%$$

$$B = \frac{\sum iVol \text{ Médio Clientes Médio Porte}}{Volume Total} \cong 15\%$$

$$C = \frac{\sum iVol \text{ Médio Clientes Pequeno Porte}}{Volume Total} \cong 5\%$$

Design do projeto

O objetivo do projeto é fornecer qual será a linha de corte para a segmentação dos clientes, a partir disto será criado um modelo de aprendizagem supervisionada que nos diga automaticamente qual segmento o cliente pertence.

Para atingir este objetivo primeiramente os dados serão trabalhados para que fiquem no formato adequado para as análises. Depois desta organização iremos classificar os dados em teste e treino para eu seja possível calibrar o modelo e realizar os ajustes necessários para constatar a confiabilidade do algoritmo que será desenvolvido.

Os clientes serão classificados por ordem crescente de Volume para que seja possível fazer a clusterização deles. Para fazer esta divisão utilizaremos o algoritmo de K-means fazendo a divisão dos clientes em 3 segmentos: grande, médio e pequeno. Cada um deles seguindo como base a teoria de Pareto (ABC).

Posteriormente será elaborado 2 modelos de classificação dos clientes com base em seu volume, um utilizando a ferramenta de Florestas Aleatórias e no outro Regressão logística. De modo que os futuros clientes já sejam classificados no segmento correto automaticamente, sem a interferência do vendedor.

Por último, os clientes serão redistribuídos entre os vendedores levando em consideração sua nova classificação com o objetivo de diminuir a quantidade de vendedores existentes.

Bibliografia

https://pt.wikipedia.org/wiki/Curva_ABC

<http://www.leg.ufpr.br/~eferreira/CE064/work5.html>

<https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd>

<https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>

<http://minerandodados.com.br/index.php/2017/12/12/entenda-o-algoritmo-k-means/>