

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Antonio C. da Silva Júnior

Título

**Curitiba
2020**

Antonio C. da Silva Júnior

Título

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. WALMES M. ZEVIANI

Curitiba
2020

Aplicação da Regressão Logística para Predição de Churn

On Dynamic of charged particles in Electric and Magnetic statics fields

Antonio C. da Silva Júnior¹

¹Campus Santos, Universidade Paulista Av. Conselheiro Nébias 766, Boqueirão, 11045-002, Santos, SP, Brasil*

Um desenvolvimento didático para determinar a solução da equação de movimento para uma partícula carregada imersa em uma região na presença de campos elétrico e magnéticos estáticos genéricos é proposto. Nossa proposta tem como alicerce a vantagem de, utilizando as propriedades da transformada de Laplace, podermos mapear um sistema de equações diferenciais não-homogêneas de segunda ordem no problema simples de encontrar as soluções de um sistema linear de equações. A partir da solução mais geral possível para o sistema, estudamos alguns casos particulares e recuperamos de forma simples alguns resultados já existentes na literatura. A fim de motivar nosso estudo, partimos do Teorema de Ehrenfest e discutimos como os resultados obtidos para o caso clássico podem ser interpretados na sua versão quântica.

Palavras-chave: movimento de cargas, campos elétrico e magnético, trajetória, análogo quântico

A didactic development to determinate the solution of the motion equations for a charged particle under influence of electric and magnetic static fields is proposed. Our proposed uses the advantages and proprieties of the Laplace's transformation, to map a system of N non-homogeneous differential equations of second order in a system composed by N linear equations. From the solution more general for dynamics of the system, we study some particular cases to recover, of a simple way, the results present in the literature. In order to give a motivation to our study, we use the Ehrenfest's theorem and we discuss as the classical results can be interpreted in its quantum version.

Keywords: Charge moving, electric and magnetic fields, trajectory, quantum analogue

1. Introdução

A importância do relacionamento de longo prazo entre cliente e empresa é um assunto vastamente discutido na literatura. Segundo Ganesh, Arnold e Reynolds (2000), devido aos efeitos de aprendizado e à redução dos custos de manutenção, atender um cliente se torna menos dispendioso a cada ano adicional de relacionamento. Para Hennig-Thurau (2004), devido ao aumento dos custos para atração de novos clientes em um mercado competitivo e à potencial redução dos custos associados aos relacionamentos de longo prazo, a retenção de clientes se torna essencial para a sobrevivência e o sucesso econômico das empresas do setor de serviços. De acordo com Gallo (2014), dependendo do estudo e do segmento no qual a empresa está inserida, o custo para adquirir um novo cliente pode ser de cinco a vinte e cinco vezes superior ao da manutenção de um cliente já existente. O desenvolvimento

de estratégias para retenção de clientes se tornou uma prática comum entre empresas de diversos segmentos, e em consequência, antever o abandono de clientes passou a ser um anseio constante. Em um momento de concentração generalizada de esforços na direção da orientação a dados, os modelos preditivos para detecção de abandono de clientes, predominantemente utilizados por grandes companhias no setor de telecomunicações, se tornaram ferramentas populares nas empresas, independentemente da magnitude e da área de atuação. A literatura comprova que a modelagem preditiva de abandono de clientes, também conhecida como modelagem preditiva de churn, é um tema bastante explorado e que possibilita inúmeras maneiras de desfecho: Botelho e Tostes (2010) ajustaram um modelo de regressão logística para prever a probabilidade de churn em uma grande empresa de varejo; Vafeiadis et al. (2015) tiveram sucesso, entre os métodos comparados, na classificação de churn através do SVM (kernel polinomial) com AdaBoost em uma empresa de telecomunicações; Com base nos dados de avaliações

*juniorssz@gmail.com

online de clientes, Kumar e Yadav (2020) propuseram um modelo preditivo de churn baseado em regras através de redes neurais artificiais e teoria dos conjuntos aproximados. Com base nos dados de uma startup brasileira que tem como principal produto uma plataforma digital para conectar vendedores de diversos segmentos aos grandes marketplaces, a proposta deste artigo é apresentar um modelo preditivo que possibilite não só a classificação de vendedores propensos a abandonar a empresa, mas que também permita a interpretação dos motivos que possivelmente estejam impactando a predição. Diante da variedade de técnicas disponíveis e das particularidades de cada modelo de negócio, a escolha do algoritmo adequado se torna uma etapa crucial do processo de modelagem. Tendo como referência a abordagem de Silva Junior, Almeida e Santos (2020), que utilizaram uma modelagem híbrida multicritério considerando múltiplos decisores para a escolha de um modelo preditivo de churn, o algoritmo escolhido para desenvolver o classificador proposto foi a regressão logística.

2. Materiais e métodos

2.1. Estruturação do conjunto de dados

Os dados utilizados neste trabalho referem-se a clientes de uma startup paranaense, anonimizados e com variáveis quantitativas padronizadas com média 0 e desvio padrão 1. Considerando que estes clientes contrataram uma plataforma digital que possibilita a venda de produtos nos principais marketplaces, neste trabalho eles serão chamados de vendedores. Devido às características da arquitetura do banco de dados e às particularidades do negócio da companhia, houve a necessidade de realizar um longo processo de data wrangling. De acordo com Kandel et al. (2011), este processo inicia-se por um diagnóstico preliminar dos dados, ou seja, se estão no formato adequado, se respondem as perguntas que motivaram a análise e o que é necessário para colocá-los no formato ideal. Em seguida avalia-se a ocorrência de dados faltantes, valores inconsistentes e duplicatas e, por fim, realiza-se um processo de limpeza e transformação, de modo a se obter um conjunto de dados adequado para o estudo.

2.1.1. Definição da resposta e covariáveis de desempenho

Inicialmente foram definidos como churn = 1 os vendedores que estiveram inativos por 30 dias corridos

Tabela 1: Definição da data de corte

Vendedor	Data de corte
Definido como churn	Última atividade
Em atividade normal	Realização da análise

Tabela 2: Interpretação das métricas de desempenho

Valor	Desempenho
0,5	Mantido
> 0,5	Aumentado
< 0,5	Reduzido

desde a data da última atividade e permaneceram no mesmo estado definitivamente, considerando como atividade o acesso à plataforma digital ou a ocorrência de uma venda online. Em seguida, em função da data de corte estabelecida conforme a Tabela 1, foram mantidos no conjunto de dados somente os vendedores com pelo menos 90 dias de histórico.

O período de 90 dias, finalizado na data de corte, foi dividido igualmente em dois quadrantes, nomeados Q1 e Q2, onde foram calculadas métricas como faturamento, ticket médio, quantidade de produtos publicados, quantidade de pedidos em cancelados, número de dias com pelo menos 1 venda e etc., em cada um dos quadrantes. Em seguida, através da Equação (1), foi calculado desempenho do vendedor em função de diversas métricas, onde $V1$ e $V2$ são os valores calculados em Q1 e Q2.

$$Desempenho = \frac{V2}{V1 + V2} \quad (1)$$

As métricas desempenho, data à natureza da equação de origem, possuem o comportamento explicado pela Tabela 2.

Referências

BOTELHO, D.; TOSTES, F. D. Modelagem de probabilidade de churn. **Revista de Administração de Empresas**, v. 50, n. 4, p. 396–410, dez. 2010.

GALLO, A. **The Value of Keeping the Right Customers**. Disponível em: <<https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>>. Acesso em: 20 jul. 2020.

GANESH, J.; ARNOLD, M. J.; REYNOLDS, K. E. Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers

and Stayers. **Journal of Marketing**, v. 64, n. 3, p. 65–87, jul. 2000.

HENNIG-THURAU, T. Customer orientation of service employees. **International Journal of Service Industry Management**, v. 15, n. 5, p. 460–478, dez. 2004.

KANDEL, S. et al. Research directions in data wrangling: Visualizations and transformations for usable and credible data. **Information Visualization**, v. 10, n. 4, p. 271–288, set. 2011.

KUMAR, H.; YADAV, R. K. Rule-Based Customer Churn Prediction Model Using Artificial Neural Network Based and Rough Set Theory. In: **Advances in Intelligent Systems and Computing**. [s.l.] Springer Singapore, 2020. p. 97–108.

SILVA JUNIOR, A. C. DA; ALMEIDA, I. D. P. DE; SANTOS, M. DOS. **Ordenação de algoritmos para modelagem preditiva de Churn: analisando o problema a partir dos métodos SAPEVO-M e VIKOR**. Congresso Internacional de Administração. **Anais...** out. 2020

VAFEIADIS, T. et al. A comparison of machine learning techniques for customer churn prediction. **Simulation Modelling Practice and Theory**, v. 55, p. 1–9, jun. 2015.