Processamento de Linguagem Natural

Atividade de Avaliação Intermediária

Atividade INDIVIDUAL

Entrega: dia da AI

Data da revisão:

Entregáveis: Jupyter notebook contendo o código completo da solução dos exercícios e trechos em markdown explicando o que foi feito.

*Objetivo da atividade: aplicar técnicas de NLP para extrair informação de uma grande coleção de textos*

Nesta tarefa, analisaremos um banco de dados de comentários de clientes, disponibilizado por uma empresa de varejo. O cliente pode fornecer feedback quanto à compra e um “rating” em estrelas.

O que gostaríamos de descobrir é:

1. Quais são os fatores *acionáveis pela empresa* que tipicamente levam clientes a darem avaliações ruins em compras? (por exemplo: saber que os clientes não gostam de fazer compras em dias de chuva não é um fator acionável; saber que estão insatisfeitos com a usabilidade do website é um fator acionável)
2. Quais são as evidências que levantamos que indicam isso?

**ENTREGÁVEIS**

Esta atividade deverá ser entregue na forma de um único Jupyter Notebook, através do Blackboard. O notebook deve, obrigatoriamente:

1. Executar sequencialmente (uma vez que o professor vai executar todos os notebooks), sem erros, assumindo as versões e APIs das bibliotecas que usamos nas aulas.
2. Usar o texto em Markdown para explicar claramente todo o raciocínio que foi realizado, acompanhando os códigos em Python. Em especial, devemos ter seções para:
   1. Introdução: com a descrição do problema e a descrição do banco de dados
   2. Desenvolvimento: com uma discussão de todas as hipóteses que foram levantadas, cada uma em sua sub-seção.
      1. Cada sub-seção deve ter um título que remeta imediatamente à hipótese que foi testada (por exemplo: “## Clientes insatisfeitos escrevem reviews mais longos?” <- não faça esse teste, porque o tamanho do review não é acionável pela empresa).
      2. Cada teste de hipótese deve culminar em uma figura que mostre os dados e a conclusão a ser tirada do teste.
      3. Faça (e mostre!) etapas de análise exploratória.
      4. Não se esqueça de evidenciar eventuais processos de limpeza dos dados.
   3. Conclusão: retomando os resultados e evidenciando os fatores que você encontrou que são mais relevantes para a satisfação do cliente.

**AVALIAÇÃO**

Os trabalhos serão avaliados de acordo com os seguintes aspectos:

1. (2 pontos) Levantamento de hipóteses: todas as hipóteses relevantes foram testadas, de acordo com as análises exploratórias? Há hipóteses que não fazem sentido para o problema (exemplo: “clientes insatisfeitos viajam para Marte?”)?
2. (2 pontos) Teste das hipóteses: usamos as metodologias corretas? As figuras estão claras? Os códigos estão corretos?
3. (2 pontos) Conclusões: elas estão baseadas nos dados gerados?
4. (2 pontos) Apresentação (texto): os textos estão claros, organizados e concisos? Eles poderiam ser lidos por uma pessoa que não entende de matemática ou de computação (dica: tente escrever o texto para uma pessoa que acabou de ingressar no primeiro ano do seu curso – pense em você no passado)
5. (2 pontos) Apresentação (figuras): os tipos de figuras usados estão adequados à mensagem que vai ser passada? Os eixos têm rótulos claros? O título das figuras guia o leitor em sua interpretação? As cores são consistentes entre as figuras do trabalho (por exemplo: “vermelho” sempre significa “cliente insatisfeito”)? A figura poderia ser destacada do trabalho escrito e, ainda assim, faria sentido?
6. (2 pontos) (bônus para quem fizer – a nota do trabalho ainda é, no máximo, 10!) É possível confirmar/desconfirmar o que você encontrou com a análise de textos através da análise de dados objetivos e numéricos também presentes no banco de dados?

**ANOTAÇÕES IMPORTANTES**

* No dia da revisão para a prova, faremos uma rodada de revisão em pares de textos/figuras em sala. Seu trabalho deve estar completo já nesse dia.
* Este trabalho é INDIVIDUAL, ORIGINAL e AUTORAL. Isso significa que você **não pode** reaproveitar o trabalho de outra disciplina, nem copiar da Internet, nem fazer em grupo.
* Você pode:
  + Encontrar soluções para pequenos problemas de programação usando o Google (por exemplo: “como faço para criar tuplas em Python”) – as documentações existem para serem usadas. Você **não pode** copiar deliberadamente a atividade-fim deste exercício.
  + Discutir com seus colegas, em alto nível, quanto a métodos e resultados que tenha tido. Você **não pode** simplesmente copiar o código ou as frases deles.
  + Procurar o Tiago (pode ser via e-mail – [tiagoft1@insper.edu.br](mailto:tiagoft1@insper.edu.br), nos horários de atendimento, ou passando diretamente na minha sala - TI-25, sétimo andar do prédio 1) para discutir dúvidas metodológicas. Você **não pode** ter uma antecipação da avaliação ou da nota.
* A entrega final é no dia da avaliação intermediária, até o horário em que a aula terminaria. Depois disso, o sistema fechará automaticamente. Não haverá extensão desse prazo uma vez que eu (Tiago) preciso corrigir as avaliações e entregar as notas e o feedback até uma data pré-definida pela Secretaria de Graduação.