

CURSO DE CIÊNCIA DE DADOS APLICADA AO PODER JUDICIÁRIO



MINERAÇÃO DE TEXTO

FICHA DO CURSO

Curso: Mineração de texto

Modalidade: EAD - autoinstrucional

Carga horária total: 16 horas

Carga horária semanal: 4 horas

Início da oferta:

Fim da oferta:

Pré-requisito: Recomenda-se que o cursista tenha conhecimentos em R e Machine learning.

Conteudistas: Walmes Marques Zeviani

1. Objetivos

Habilitar o(a) aluno(a) a utilizar o software R de computação estatística e gráficos para realizar análise com dados na forma de texto em linguagem natural.

Ao final do curso o participante será capaz de configurar o ambiente para análise de texto no R, importar e pré-processar corpus de texto, realizar visualizações de dados e aplicar métodos como análise de sentimento, modelagem de tópicos, análise de agrupamento, rotulação de partes do discurso e word embeddings.

2. Programa

Problemas e métodos em mineração de textos. Importação e pré-processamento de texto. Análise de sentimentos. Tokenização e n-gramas. Frequência de palavras. Nuvem de palavras. Matrizes de documentos e termos e ponderações. Análise de agrupamento. Modelagem de tópicos. Alocação latente de Dirichlet. Aplicações de NLP. Rotulação de partes do discurso. Word embeddings.

3. Procedimentos didáticos

A cada módulo (em um total de **4 módulos**) serão desenvolvidas as seguintes atividades pelos cursistas:

1. Acessar o material pré aula para preparar-se para a aula.
2. Assistir as vídeos-aulas programadas para o módulo
3. Fazer o estudo individual de materiais indicados como: leituras complementares, resolução de exercícios e acesso a vídeos adicionais.
4. Realizar as atividades de avaliação do módulo no formato de quiz.

Como suporte ao Ensino Remoto será utilizada a infraestrutura da plataforma moodle do Centro de Formação e Aperfeiçoamento de Servidores do Poder Judiciário (CEAJUD).

4. Procedimentos metodológicos

Conjunto de vídeo-aulas separadas por módulos de conhecimento acompanhado de avaliações para reforçar o aprendizado.

5. Distribuição da carga horária do cursista

A Tabela 1 a seguir apresenta distribuição da carga horária semanal total do cursista, de 4 horas por semana, nas atividades semanais previstas na vigência do curso.

Tabela 1 - Carga horária semanal do(a) cursista.

Atividade semanal	Carga horária (horas)	Fração (%)
Estudo individual de conteúdo pré-aula	00:30	12,5%
Assistir vídeo-aulas	02:00	50%
Estudo individual pós aula	00:30	12,5%
Atividades avaliativas	01:00	25%
Total	04:00	100%

6. Conteúdo programado

O curso está dividido em 4 módulos sendo que cada módulo configura uma semana de atividades. O conteúdo programado é apresentado na Tabela 2.

Tabela 2 - Cronograma detalhado do conteúdo das unidades didáticas.

Módulo	Conteúdo programado	Duração
1	Problemas e métodos em mineração de textos. Importação e pré-processamento de texto.	4 horas
2	Matrizes de documentos e termos e ponderações. Análise de sentimentos.	4 horas
3	Modelo de linguagem e n-gramas. Análise de agrupamento de documentos. Modelagem de tópicos.	4 horas
4	Word embeddings. Aplicações de NLP.	4 horas
Encerramento do curso		

7. Desempenho no curso

O desempenho no curso será determinado pela nota em atividade avaliativa aplicada aos cursistas. Requer-se, no mínimo, 70% de aproveitamento para obtenção de certificado.

8. Referências bibliográficas

1. Silge, Julia, and David Robinson. **Text mining with R : a tidy approach**. Sebastopol, CA: O'Reilly Media, 2017. Print.

2. Kumar, Ashish, and Avinash Paul. **Mastering text mining with R : master text-taming techniques and build effective text-processing applications with R**. Birmingham, UK: Packt Publishing, 2016. Print.
3. Kwartler, Ted. **Text mining in practice with R**. Hoboken, NJ: John Wiley & Sons, 2017. Print.
4. Danneman, Nathan, Richard Heimann, and Wood. **Social media mining with R : deploy cutting-edge sentiment analysis techniques to real-world social media data using R**. Birmingham, England: Packt Publishing, 2014. Print.
5. Miner, Gary. **Practical text mining and statistical analysis for non-structured text data applications**. Amsterdam: Academic Press, 2012. Print.

9. Informações sobre os conteudistas

Walmes Marques Zeviani

Engenheiro Agrônomo pela Universidade Federal da Grande Dourados (2003-2007) e Mestre (2008-09) e Doutor (2009-13) em Estatística & Experimentação Agropecuária pela Universidade Federal de Lavras. Professor Adjunto III do Departamento de Estatística da Universidade Federal do Paraná (2010-hoje). Atua principalmente com ensino no Curso de Bacharel em Estatística e na Especialização em Data Science & Big Data (UFPR). Tem experiência com cursos de capacitação do software R para planejamento e análise de experimentos, modelos de regressão linear e não linear, manipulação e visualização de dados e construção de interfaces gráficas e pacotes R.

10. Requisitos Técnicos

Computador com acesso à internet. Permissão para instalar programas.

CURSO DE CIÊNCIA
DE DADOS APLICADA
AO PODER JUDICIÁRIO

