

MINERAÇÃO DE TEXTO

FICHA DO CURSO

1. Resumo do curso

Curso: Mineração de texto

Modalidade: Ensino à distância

Carga horária total: 16 horas

Carga horária síncrona: 2,5 horas por semana

Início da oferta: Agosto/2022

Pré-requisito: R. Machine learning.

Duração do curso: 4 semanas

Carga horária semanal: 4 horas

Carga horária assíncrona: 1,5 horas por semana

Fim da oferta: Setembro/2022

2. 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 de extração de entidades nomeadas.

3. 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. Métodos alternativos. Aplicações de NLP. Rotulação de partes do discurso. Extração de entidades nomeadas. Word embeddings.

4. Procedimentos didáticos

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

1. Acessar o material pré aula para preparar-se para a aula.
2. Assistir as aulas programadas para a semana no horário definido.
3. Trabalhar nos tutoriais guiados disponibilizados pelo professor.
4. Fazer o estudo individual de materiais indicados como: leituras complementares, resolução de exercícios e acesso a vídeos adicionais.
5. Participar do fórum do curso contribuindo com tópicos para a discussão ou respondendo e complementando tópicos em aberto relacionados ao conteúdo apresentado.
6. Realizar as atividades de avaliação semanal de desempenho e avaliação do curso.

Como suporte ao Ensino Remoto será utilizada a infraestrutura de AVA (Ambiente Virtual de Aprendizado) que é acessado pelo endereço <https://treinamento.omegadatascience.com.br> para cursistas cadastrados.

Serão observados os seguintes detalhes sobre os procedimentos didáticos no período de vigência do curso, conforme itens a seguir.

4.1 Sistema de comunicação durante aulas ao vivo

Será usado o sistema de comunicação por mensagens instantâneas do serviço de streaming utilizado para transmissão das aulas ao vivo durante a apresentação do professor. Durante a apresentação do professor, a comunicação por texto é com a equipe de monitores do curso. Durante a aula, em momentos de pausa para dúvidas definidos pelo professor, os alunos poderão interagir por voz e vídeo.

4.2 Sistema de comunicação entre as aulas ao vivo

Será usado o recurso de fórum do AVA, permitindo que alunos abram tópicos e contribuam com a discussão ou resolução de dúvidas em tópicos abertos.

4.3 Sistema de comunicação nos plantões de dúvida

Será usado o sistema de comunicação por mensagens instantâneas, voz e vídeo do serviço de streaming utilizado para transmissão dos plantões de dúvidas.

4.4 Materiais didáticos

1. No Moodle terá o guia rápido do curso, onde constarão detalhes das plataformas de ensino, instruções sobre a comunicação com os instrutores, monitores e coordenação, cronograma e programa da disciplina, orientação das atividades remotas assíncrona e síncrona, seus prazos e pesos na determinação de desempenho e participação.
2. Dentro da conta de cada usuário na Sympla estarão disponíveis os links para a transmissão das aulas e dos plantões de dúvidas.
3. No Moodle estarão os links para acesso às transmissões gravadas das aulas. Elas ficarão hospedadas no canal do Youtube designado para o treinamento.
4. Slides (em formato pdf) usados nas aulas bem como eventuais arquivos como scripts, planilhas e arquivos de dados serão disponibilizados no Moodle.
5. No Moodle será usado o recurso de fórum para cursistas abrirem e contribuírem em tópicos, tirarem dúvidas e contribuírem com materiais.
6. No Moodle poderão ser indicados materiais complementares àqueles produzidos pelos professores do curso, como: vídeos que demonstram conceitos e ideias sobre os conteúdos das unidades didáticas e links para outros sítios eletrônicos contendo materiais relacionados aos temas das aulas.
7. No Moodle serão disponibilizadas atividades avaliativas de desempenho e participação bem como formulários de avaliação da qualidade do curso.

4.5 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
Participação na aula ao vivo	01:30	37,5
Estudo individual pós aula	00:30	12,5
Participação no plantão de dúvidas	01:00	25,0
Participação no fórum	00:30	12,5
Total	04:00	100

4.6 Modelo de tutoria a distância

A dinâmica deste curso é disponibilizar para o cursista um serviço de autoaprendizagem que consta de vídeo-aulas síncronas com acesso posterior às gravações, disponibilização de materiais didáticos usados nos vídeos e complementares, atividades para a prática do conteúdo.

Além das aulas e respectivos materiais, o cursista terá um encontro remoto síncrono de plantão de dúvidas para interagir com o professor e pares.

No AVA haverá durante todo período de vigência do curso, fórum separados por encontro no qual os alunos poderão abrir e contribuir em tópicos diversos, como por exemplo, dúvidas em relação ao conteúdo, solicitar e/ou recomendar materiais complementares, contribuir com tópicos abertos pelos pares, apresentar resultados de atividades desenvolvidas, entre muitas outras possibilidades. Será gerado um FAQ (frequently asked questions) com as questões mais relevantes ou populares do fórum.

Em cada unidade didática, o cursista poderá avaliar seu aprendizado por meio de quizzes sobre os tópicos cobertos no encontro imediatamente anterior.

E por fim, toda unidade didática terá uma enquete anônima com questões voltadas à avaliação do curso e espaço de texto livre para expressar sua opinião.

5 Conteúdo programado

O curso está dividido em unidades didáticas (UD), sendo uma UD para cada semana de atividades. Na sequência, estão listadas as 10 UD's, com respectivo conteúdo planejado, na Tabela 2.

Tabela 2 - Cronograma detalhado do conteúdo das unidades didáticas.

UD	Conteúdo previsto	Período
1	Problemas e métodos em mineração de textos. Importação e pré-processamento de texto. Análise de sentimentos.	29/08 - 02/09
2	Tokenização e n-gramas. Frequência de palavras. Nuvem de palavras. Matrizes de documentos e termos e ponderações. Análise de agrupamento.	05/09 - 09/09

3	Modelagem de tópicos. Alocação latente de Dirichlet. Métodos alternativos.	12/09 - 16/09
4	Aplicações de NLP. Rotulação de partes do discurso. Extração de entidades nomeadas. Word embeddings.	19/09 - 23/09
Encerramento do curso		

6 Avaliação de desempenho e participação

6.1 Desempenho no curso

O desempenho no curso será determinado pela nota média das 3 maiores notas dos 4 quizzes semanais aplicadas aos cursistas.

6.2 Participação no curso

A frequência de participação no curso será determinada pela realização dos quizzes semanais bem como pelo preenchimento das enquetes semanais de avaliação do curso ou questões do material pré-aula/pós-aula. Receberão certificados de participação aqueles com 70% ou mais de aproveitamento nestas atividades.

7 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.

8 Informações sobre o professor

Abaixo estão as informações dos membros da equipe envolvidos diretamente com os cursistas.

Tabela 3 – Informações dos membros da equipe.

Função	Nome	Contato
Instrutores do curso	Walmes Marques Zeviani	walmes@ufpr.br
Monitor do curso	A definir	A definir
Coordenador	Walmes Marques Zeviani	walmes@ufpr.br

CURSO DE CIÊNCIA
DE DADOS APLICADA
AO PODER JUDICIÁRIO

