



# UNIVERSIDADE FEDERAL DE SANTA CATARINA

## PROGRAMA DE DISCIPLINA

DADOS DA DISCIPLINA
Tipo de disciplina a propor: <input checked="" type="checkbox"/> Normal <input type="checkbox"/> Estágio Docência <input type="checkbox"/> Tese/Dissertação
Nome: <b>Mineração de Texto Aplicada</b>
Periodicidade de oferecimento da disciplina: <input checked="" type="checkbox"/> Anual <input type="checkbox"/> Semestral <input type="checkbox"/> Esporádica
Ano/Período a ser oferecida pela primeira vez: 2021.2

CARGA HORÁRIA		
Total de horas-aula: 60	Total de créditos: 4	
Especifique como será distribuída a carga horária da disciplina conforme orientado abaixo:		
Carga horária teórica: (1 crédito = CH 15)	Carga horária prática: (1 crédito = CH 30)	Carga horária teórico-prática: (1 crédito = CH 30)
60		
Alteração de disciplina: <input type="checkbox"/> Não <input checked="" type="checkbox"/> Sim: Código ou nome da disciplina a ser alterada: <b>Mineração de Texto Aplicada</b>		
Nível a ser oferecida: <input type="checkbox"/> Mestrado <input type="checkbox"/> Doutorado <input checked="" type="checkbox"/> Mestrado e Doutorado		
Tipo de Disciplina		
Mestrado: <input checked="" type="checkbox"/> Eletiva - <input type="checkbox"/> Obrigatória	Doutorado: <input checked="" type="checkbox"/> Eletiva - <input type="checkbox"/> Obrigatória	

Corpo Docente Responsável
Moisés Lima Dutra

Ementa:
Técnicas de mineração de texto. Corpora textuais. Web <i>scraping</i> . Higienização e normalização textual. Processamento de linguagem natural. Técnicas de <i>machine learning</i> para textos. Avaliação e visualização de resultados.

ÁREA DE CONCENTRAÇÃO
Mestrado: Gestão da Informação
Doutorado: Gestão da Informação

METODOLOGIA
Aulas expositivas, seminários, discussões, e concepção de um projeto de mineração textual. O Moodle ( <a href="http://moodle.ufsc.br/">http://moodle.ufsc.br/</a> ) será utilizado como instrumento de suporte da disciplina. Portanto, o mesmo deve ser consultado todas as semanas.

FORMA DE AVALIAÇÃO
Composição da nota: <ul style="list-style-type: none"><li>• Apresentação de seminário temático e participação nas discussões: 30%;</li><li>• Projeto de mineração textual: 70%.</li></ul>



# UNIVERSIDADE FEDERAL DE SANTA CATARINA

## BIBLIOGRAFIA BÁSICA

BENGFORT, Benjamin; BILBRO, Rebecca; OJEDA, Tony. **Applied Text Analysis with Python: Enabling Language Aware Data Products with Machine Learning**. Sebastopol, Ca: O'reilly, 2017.

INGERSOLL, Grant S.; MORTON, Thomas S.; FARRIS, Andrew L. **Taming Text: How to find, organize and manipulate it**. Shelter Island, NY (USA): Manning Publications Co., 2013. 298 p.

SARKAR, Dipanjar. **Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data**. Middletown, DE (USA): Apress, 2016. 385 p.

SUMMERFIELD, Mark. **Programação em Python 3: Uma Introdução Completa à Linguagem Python**. Rio de Janeiro: Alta Books, 2013. 506 p.

WEISS, Sholom M.; INDURKHYA, Nitin; ZHANG, Tong. **Fundamentals of Predictive Text Mining**. New York: Springer, 2010. 226 p. (Texts in Computer Science).

## JUSTIFICATIVA DO DOCENTE PARA A OFERTA DA DISCIPLINA

A moderna pesquisa científica baseia-se cada vez mais no processamento de dados e informações gerados em meios digitais. Todavia, apesar da crescente adoção de formatos abertos e estruturados, a maioria absoluta dos dados e informações existentes em meios digitais não estão disponíveis em formatos pré-definidos, mas sim de maneira completamente desestruturada, ou seja, em formato de texto. Mineração de Texto é um processo que busca extrair informações de alta qualidade – que se referem a uma combinação de relevância, novidade e interesse –, que são extraídas com o suporte da identificação de padrões e tendências, por meio da aplicação de métodos advindos da Estatística e da Inteligência Artificial, mais especificamente, do Aprendizado de Máquina (*Machine Learning*). A Mineração de Texto envolve a coleta automática de dados textuais, seguida de sua estruturação, análise, adição e/ou retirada de elementos linguísticos, transformação e representação de texto como números, aplicação de técnicas estatísticas e de aprendizado de máquina, avaliação e interpretação dos resultados. Estudar este contexto tecnológico por meio do desenvolvimento e aplicação de técnicas de mineração de texto configura-se, portanto, como uma necessidade corrente e atual. O estudo dos tópicos desta disciplina ajudará o estudante a compreender, desenvolver e utilizar a mineração de texto como ferramenta de suporte para suas pesquisas acadêmicas e para sua vida profissional.

Data: 10/06/2021

---

Assinatura do Docente  
Responsável pela disciplina



# UNIVERSIDADE FEDERAL DE SANTA CATARINA

## DELIBERAÇÃO DO COLEGIADO PLENO

( ) Não Aprovado ( ) Aprovado

Nível para oferta: ( ) Mestrado ( ) Doutorado ( ) Mestrado e Doutorado

Data da Reunião: \_\_\_\_/\_\_\_\_/\_\_\_\_

**Encaminhe-se à CÂMARA DE PÓS-GRADUAÇÃO**

Assinatura do Presidente do Colegiado Pleno

## DELIBERAÇÃO DA CÂMARA DE PÓS-GRADUAÇÃO

( ) Não Aprovado ( ) Aprovado

Nível para oferta: ( ) Mestrado ( ) Doutorado ( ) Mestrado e Doutorado

Data da Reunião: \_\_\_\_/\_\_\_\_/\_\_\_\_ - Conforme folhas \_\_\_\_ do processo.

## CRIAÇÃO DA DISCIPLINA (Registro no CAPG)

Nome da Disciplina:

Código:

Carga horária:

Hora-aula semanal: \_\_\_\_ créditos | Número de crédito total: \_\_\_\_ horas/aula

Carga horária teórica:

Carga horária prática:

Carga horária teórico-prática:

Nível: ( ) Mestrado ( ) Doutorado ( ) Mestrado e Doutorado

Tipo de disciplina:

Registro no sistema feito em: \_\_\_\_/\_\_\_\_/\_\_\_\_ Por: \_\_\_\_\_