

RENCANA PEMBELAJARAN SEMESTER PROGRAM STUDI SARJANA TERAPAN TEKNIK INFORMATIKA DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER									
Kode		Mata Kuliah	Text Mining					NoId: RF-DTEL-PSTI-4.05.Rev.01[031]	
Bobot SKS	2	Semester	3						
Kelompok MK	MK inti prodi	Jam/minggu	2						
Tim Pengampu MK	Tita Karlita / M. Rosyid Mubtadai								
Capaian Pembelajaran	<p>Mahasiswa mampu menyelesaikan masalah – masalah menggunakan metode text mining yang tepat berdasarkan supervised dan unsupervised baik secara individu maupun berkelompok/kerjasama tim. Capaian pembelajaran secara detail adalah sebagai berikut:</p> <ul style="list-style-type: none"> • KK3 Keahlian batch processing atau real-time processing pada data yang dikumpulkan dan disimpan • KK5 Keahlian ekstraksi makna dan interpretasi data, menggunakan berbagai alat dan metode pembelajaran matematika dan mesin • KK24 Keahlian menyaring gagasan, sistem dan data yang kompleks menjadi informasi yang dapat diterapkan • KK28 Kemampuan analitis untuk membaca dan memahami data, kreativitas untuk menciptakan nilai menggunakan data dan kekuatan bisnis untuk berinovasi dengan memperluas nilai. • P28 Menguasai pengetahuan menerapkan prinsip-prinsip ilmu data, dan alat dan teknik statistik dan analitis, dalam bidang ilmiah, sosial dan industri yang kompleks. 								
Pokok Bahasan	<ol style="list-style-type: none"> 1. Introduction to Text Mining dan Natural Language Processing 2. Data Preparation 3. Bag-of-Words 4. Word Embeddings 5. Text Classification 6. Image Captioning 7. Text Summarization 								
Referensi	<ol style="list-style-type: none"> 1. Deep Learning for Natural Language Processing, Jason Brownlee 2. <u>Natural Language Processing with Python</u>, Steven Bird, Ewan Klein and Edward Loper 3. Taming Text, <u>Grant Ingersoll</u>, <u>Thomas Morton</u> and <u>Drew Farris</u>. 4. Text Mining with R, Julia Silge and David Robinson. 5. Foundations of Statistical Natural Language Processing, <u>Christopher Manning</u> and Hinrich Schütze. 6. <u>Speech and Language Processing</u>, Daniel Jurafsky and James Martin. 7. Statistical Machine Translation, Philipp Koehn. 8. <u>Statistical Methods for Speech Recognition</u>, Frederick Jelinek. 								
MK Prasyarat	Text processing, Artificial Intelligence								
Media Pembelajaran	Software: OS Windows, Python dll Hardware: PC/Laptop, LCD Projector								
Asesmen (%)	Praktek (50 %), Tugas (40 %), Sikap (10 %)								
Mgg Ke-	Sub Capaian Pembelajaran MK (Kemampuan Akhir Yang Direncanakan)	Bahan Kajian (Materi Pembelajaran)	Bentuk Pembelajaran	Waktu Belajar (menit)	Kriteria Asesmen (Indikator)	Bentuk Asesmen	Bobot		

(1)	Memahami text mining, beda data mining dan text mining, beda data mining dan data analytics, beda data mining dan natural language processing	<ul style="list-style-type: none"> • Pengenalan text mining • Mengetahui apakah Text Mining • Mengetahui alasan menggunakan Text Mining • Memahami beda data mining dan text mining, beda data mining dan data analytics, beda data mining dan natural language processing 	Kuliah Pengantar & Brainstorming, Diskusi	TM: 100 menit Tgs: 100 menit BM: 100 menit			5%
(2)	<p>Data Preparation 1:</p> <ul style="list-style-type: none"> • How to Clean Text Manually and with NLTK • How to Prepare Text Data with scikit-learn 	<ul style="list-style-type: none"> • How to Clean Text Manually and with NLTK <ul style="list-style-type: none"> a. How to get started by developing your own very simple text cleaning tools. b. How to take a step up and use the more sophisticated methods in the NLTK library. c. Considerations when preparing text for natural language processing models. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Metamorphosis by Franz Kafka 2. Text Cleaning is Task Specific 3. Manual Tokenization 4. Tokenization and Cleaning with NLTK 5. Additional Text Cleaning Considerations <ul style="list-style-type: none"> • How to Prepare Text Data with scikit-learn <ul style="list-style-type: none"> a. How to convert text to word count vectors with CountVectorizer. b. How to convert text to word frequency vectors with TfidfVectorizer. c. How to convert text to unique integers with HashingVectorizer. • How to Prepare Text Data With Keras <ul style="list-style-type: none"> a. About the convenience methods that you can use to quickly prepare text data. b. The Tokenizer API that can be used on training data and used to encode 	Kuliah, Diskusi kelompok, Praktek, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit	o		10%

		<p>training, validation, and test documents.</p> <p>c. The range of 4 different document encoding schemes offered by the Tokenizer API.</p> <p>Tutorial Overview</p> <ol style="list-style-type: none"> 1. Split words with text to word sequence. 2. Encoding with one hot. 3. Hash Encoding with hashing trick. 4. Tokenizer API 					
(3)	Bag-of-Words 1: The Bag-of-Words Model	<ul style="list-style-type: none"> • What the Bag-of-words model is and why it is needed to represent text. • How to Develop a bag-of-words model for a collection of documents. • How to Use Different Techniques to prepare a vocabulary and score words. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. The Problem with Text 2. What is a Bag-of-Words? 3. Example of the Bag-of-Words Model 4. Managing Vocabulary 5. Scoring Words 6. Limitations of Bag-of-Words 	Kuliah, Diskusi kelompok, Praktek, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit	○		5%
(4)	Bag-of-Words 2: Prepare Movie Review Data for Sentiment Analysis	<ul style="list-style-type: none"> • How to Load text data and clean it to remove punctuation and other non-words. • How to Develop a vocabulary, tailor it, and save it to file. • How to Prepare movie reviews using cleaning and a predefined vocabulary and save them to new files ready for modeling. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Movie Review Dataset 2. Load Text Data 3. Clean Text Data 4. Develop Vocabulary 5. Save Prepared Data 	Kuliah, Diskusi kelompok, Praktek, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit	○		10%
(5)	Bag-of-Words 3: Project: Develop a Neural Bag-of-Words Model for Sentiment Analysis	<ul style="list-style-type: none"> • How to Prepare the review text data for modeling with a restricted vocabulary. • How to Use the bag-of-words model to prepare train and test data. 	Kuliah, Diskusi kelompok, Praktek, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit	○		10%

		<ul style="list-style-type: none"> • How to Develop a Multilayer Perceptron bag-of-words model and use it to make predictions on new review text data. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Movie Review Dataset 2. Data Preparation 3. Bag-of-Words Representation 4. Sentiment Analysis Models 5. Comparing Word Scoring Methods 6. Predicting Sentiment for New Reviews 					
(6)	<p>Word Embeddings 1:</p> <ul style="list-style-type: none"> • The Word Embedding Model (teori) • How to Develop Word Embeddings with Gensim 	<p>The Word Embedding Model (teori)</p> <ul style="list-style-type: none"> • What the word embedding approach for representing text is and how it differs from other feature extraction methods. • That there are 3 main algorithms for learning a word embedding from text data. • That you can either train a new embedding or use a pre-trained embedding on your natural language processing task. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. What Are Word Embeddings? 2. Word Embedding Algorithms 3. Using Word Embeddings <p>How to Develop Word Embeddings with Gensim</p> <ul style="list-style-type: none"> • How to train your own Word2Vec word embedding model on text data. • How to visualize a trained word embedding model using Principal Component Analysis. • How to load pre-trained Word2Vec and GloVe word embedding models from Google and Stanford. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Word Embeddings 2. Gensim Library 3. Develop Word2Vec Embedding 4. Visualize Word Embedding 5. Load Google's Word2Vec Embedding 	<p>Kuliah, Diskusi kelompok, Praktek, maju di depan kelas</p>	<p>TM: 100 menit Tgs: 100 menit BM: 100 menit</p>			5%

		6. Load Stanford's GloVe Embedding					
(7)	Word Embeddings 2:	<p>How to Learn and Load Word Embeddings in Keras</p> <ul style="list-style-type: none"> • About word embeddings and that Keras supports word embeddings via the Embedding layer. • How to learn a word embedding while fitting a neural network. • How to use a pre-trained word embedding in a neural network. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Word Embedding 2. Keras Embedding Layer 3. Example of Learning an Embedding 4. Example of Using Pre-Trained GloVe Embedding 5. Tips for Cleaning Text for Word Embedding 	Kuliah, Diskusi kelompok, Praktek, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit			5%
UTS							
(9)	Text Classification 2: Text classification using Statistical methods	Text classification using Naïve bayes, Logistic regression, Boosting model K-Nearest Neighbours, grid search methods	Kuliah, Diskusi kelompok, Praktek, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit	○		5%
(10)	Text Classification 2: • Neural Models for Document Classification (teori) • Project: Develop an Embedding + CNN Model for Sentiment Analysis	<p>Neural Models for Document Classification (teori)</p> <ul style="list-style-type: none"> • The general combination of deep learning methods to consider when starting your text classification problems. • The first architecture to try with specific advice on how to configure hyperparameters. • That deeper networks may be the future of the eld in terms of flexibility and capability. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Word Embeddings + CNN = Text Classification 2. Use a Single Layer CNN Architecture 3. Dial in CNN Hyperparameters 4. Consider Character-Level CNNs 	Kuliah, Diskusi kelompok, Praktek, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit	○		5%

		<p>5. Consider Deeper CNNs for Classification</p> <p>Project: Develop an Embedding + CNN Model for Sentiment Analysis</p> <ul style="list-style-type: none"> • How to prepare movie review text data for classification with deep learning methods. • How to develop a neural classification model with word embedding and convolutional layers. • How to evaluate the developed a neural classification model. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Movie Review Dataset 2. Data Preparation 3. Train CNN With Embedding Layer 4. Evaluate Model 					
(11)	<p>Text Classification 3: Project: Develop an n-gram CNN Model for Sentiment Analysis</p>	<p>Project: Develop an n-gram CNN Model for Sentiment Analysis</p> <ul style="list-style-type: none"> • How to prepare movie review text data for modeling. • How to develop a multichannel convolutional neural network for text in Keras. • How to evaluate a fit model on unseen movie review data. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Movie Review Dataset. 2. Data Preparation. 3. Develop Multichannel Model. 4. Evaluate Model. 	<p>Kuliah, Diskusi kelompok, maju di depan kelas</p>	<p>TM: 100 menit Tgs: 100 menit BM: 100 menit</p>	○		10%
(12)	<p>Image Captioning 1:</p> <ul style="list-style-type: none"> • Neural Image Caption Generation (teori) • Neural Network Models for Caption Generation (teori) • How to Load and Use a Pre-Trained Object Recognition Model 	<p>Neural Image Caption Generation (teori)</p> <ul style="list-style-type: none"> • About the challenge of generating textual descriptions for images and the need to combine breakthroughs from computer vision and natural language processing. • About the elements that comprise a neural feature captioning model, namely the feature extractor and language model. • How the elements of the model can be arranged into an Encoder-Decoder, possibly with the use of an attention mechanism. 	<p>Kuliah, Diskusi kelompok, maju di depan kelas</p>	<p>TM: 100 menit Tgs: 100 menit BM: 100 menit</p>	○		10%

		<p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Describing an Image with Text 2. Neural Captioning Model 3. Encoder-Decoder Architecture <p>Neural Network Models for Caption Generation (teori)</p> <ul style="list-style-type: none"> • The challenge of caption generation and the use of the encoder-decoder architecture. • The inject model that combines the encoded image with each word to generate the next word in the caption. • The merge model that separately encodes the image and description which are decoded in order to generate the next word in the caption. <p>How to Load and Use a Pre-Trained Object Recognition Model</p> <ul style="list-style-type: none"> • About the ImageNet dataset and competition and the VGG winning models. • How to load the VGG model in Keras and summarize its structure. • How to use the loaded VGG model to classifying objects in ad hoc photographs. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. ImageNet 2. The Oxford VGG Models 3. Load the VGG Model in Keras 4. Develop a Simple Photo Classifier 					
(13)	Image Captioning 2: How to Evaluate Generated Text With the BLEU Score	<ul style="list-style-type: none"> • How to Evaluate Generated Text With the BLEU Score • A gentle introduction to the BLEU score and an intuition for what is being calculated. • How you can calculate BLEU scores in Python using the NLTK library for sentences and documents. • How you can use a suite of small examples to develop an intuition for how 	Kuliah, Diskusi kelompok, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit	o		5%

		<p>differences between a candidate and reference text impact the final BLEU score.</p> <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Bilingual Evaluation Understudy Score 2. Calculate BLEU Scores 3. Cumulative and Individual BLEU Scores 4. Worked Examples <p>How to Prepare a Photo Caption Dataset For Modeling</p> <ul style="list-style-type: none"> • About the Flickr8K dataset comprised of more than 8,000 photos and up to 5 captions for each photo. • How to generally load and prepare photo and text data for modeling with deep learning. • How to specically encode data for two di erent types of deep learning models in Keras. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Download the Flickr8K Dataset 2. How to Load Photographs 3. Pre-Calculate Photo Features 4. How to Load Descriptions 5. Prepare Description Text 6. Whole Description Sequence Model 7. Word-By-Word Model 8. Progressive Loading <p>Project: Develop a Neural Image Caption Generation Model</p> <ul style="list-style-type: none"> • How to prepare photo and text data for training a deep learning model. • How to design and train a deep learning caption generation model. • How to evaluate a train caption generation model and use it to caption entirely new photographs. <p>Tutorial Overview:</p> <ol style="list-style-type: none"> 1. Photo and Caption Dataset 					
--	--	---	--	--	--	--	--

		2. Prepare Photo Data 3. Prepare Text Data 4. Develop Deep Learning Model 5. Evaluate Model 6. Generate New Captions					
(14)	Text Summarization 1	<ul style="list-style-type: none"> • Introduction to Text Summarization • Encoder-Decoder Deep Learning Models for Text Summarization 	Kuliah, Diskusi kelompok, maju di depan kelas	TM: 100 menit Tgs: 100 menit BM: 100 menit			5%
(15)	Text Summarization 2	<ul style="list-style-type: none"> • Prepare News Articles for Text Summarization • Encoder-Decoder Models for Text Summarization in Keras 	Presentasi Proyek	TM: 100 menit Tgs: 100 menit BM: 100 menit			10%

UAS

Keterangan:
 TM : Tatap Muka
 Tgs : Tugas
 BM : Belajar Mandiri