

## 《语言数据挖掘与机器学习》课程教学大纲（2020 版）

课程基本信息 (Course Information)				
课程代码 (Course Code)		*学时 (Credits Hours )	32	*学分 (Credits)
				2
*课程名称 (Course Name)	(中文) 语言数据挖掘与机器学习			
	(英文) Data Mining & Machine Learning			
课程类型 (Course Type)	基础模块必须课			
授课对象 (Target Audience)	微专业学员			
授课语言 (Language of Instruction)	全中文			
*开课院系 (School)	外国语学院			
先修课程 (Prerequisite)	编程与语言数据分析	后续课程 (post)		
*课程负责人 (Instructor)	管新潮	课程网址 (Course Webpage)		

<p>*课程简介 (中文) (Description)</p>	<p>本课程是语言数据科学微专业应用模块课程之一，旨在培养学生将机器学习技术应用于数据挖掘与分析的能力。课程具体内容涉及：一是经典机器学习模式的支持向量机、Logistic 回归、朴素贝叶斯分类、多种聚类模式、决策树分类、降维处理、异常检测、结构化预测；二是深度学习模式的线性回归、Logistic 回归、文本分类；三是语言模型模式的跨语言模型、语料库微调、模型组合等；四是文本向量化。该课程结合了语言学、Python 编程和统计学等方面知识，以实践应用为导向，注重理论学习与实践操作的融合。</p> <p>课程以案例形式展开实际教学活动，学生可以通过具体案例学习掌握机器学习技术应用于语言学/翻译学数据外挖掘的融合模式，熟悉并了解具体技术的适用性和可应用性。通过本课程学习，学生可以掌握数据挖掘与分析的基本机器学习方法，习得机器学习相关的数据分析能力，为适应该领域的后续发展铺垫知识技术基础。</p>
<p>*课程简介 (英文) (Description)</p>	<p>"Data Mining &amp; Machine Learning" is one of the applied courses for the Micro-degree of Language Data Science, aimed at cultivating students' ability to apply different machine learning technology in the data mining and analysis. The course consists of four parts: First, classical machine learning models including Support Vector Machine, Logistic Regression, Naive Bayes Classification, various clustering models, Decision Tree Classification, dimensionality reduction, anomaly detection, and structured prediction; second, deep learning models such as Linear Regression, Logistic Regression, and text classification; third, language model approaches including cross-lingual models, corpus fine-tuning, and model ensembles; and fourth, text vectorization. These topics provide students with a comprehensive understanding of both traditional and modern machine learning approaches, as well as the latest advancements in deep learning and natural language processing.</p> <p>This course is conducted through case studies, allowing students to learn and master the integration model of machine learning techniques and linguistics/translation studies through specific cases, and understand the applicability and applicability of specific technologies. Through this course, students can master the basic methods of data mining and analysis, acquire certain abilities in machine learning, and get knowledge and technical skills for adapting to the future development in this field.</p>
<p><b>课程目标与内容 (Course objectives and contents)</b></p>	

<p>*课程目标 (Course Object)</p>	<p>结合本校办学定位、学生情况、专业人才培养要求，具体描述学习本课程后应该达到的知识、能力、素质、价值水平。</p> <p>1.能了解机器学习技术在本学科的基本应用方法，认识到技术应用有其独特的语言学意义，提升专业自信心。（A3）</p> <p>2.能以创新方式将机器学习技术应用于语言学问题的数据挖掘处理，掌握并提升解决相关问题的能力。（B1）</p> <p>3.善于发现语言学与大数据相结合所产生的新问题，并为此应用合适的自然语言处理技术，养成独立解决问题的能力。（B3）</p> <p>4.能开展机器学习技术的语言学适应性验证，以确定相应技术是否适宜于为语言学目的所用。（B2）</p> <p>5.能通过语言知识与技术的融合性学习，解决涉及语言产品的细腻性问题，适应社会对高素质语言专业人才的需求。（B5）</p>						
<p>*教学内容进度安排及对应课程目标 (Class Schedule &amp; Requirements &amp; Course Objectives)</p>	<p>章节</p>	<p>教学内容 (要点)</p>	<p>学时</p>	<p>教学形式</p>	<p>作业及考核要求</p>	<p>课程思政融入点</p>	<p>对应课程目标</p>
<p>示例:</p>							
<p>数据挖掘与机器学习概述</p>	<p>讲授课程所涉具体内容及其意义和作用</p>	<p>2</p>	<p>以讲授为主</p>	<p>文章阅读</p>	<p>认识到与语言学相结合的机器学习技术其社会价值和意义，提升专业热情</p>	<p>1,2</p>	
<p>经典机器学习技术</p>	<p>朴素贝叶斯分类器和支持向量机</p>	<p>4</p>	<p>案例讲解和讨论</p>	<p>应用于自选文本</p>	<p>认识到经典工具技术的现实意义和作用，培养严谨认真的专业态度</p>	<p>1,2</p>	
<p>文本向量化</p>	<p>文本知识表达方法</p>	<p>2</p>	<p>案例讲解和讨论</p>	<p>应用于自选文本并比较</p>	<p>认识到文本知识表达的现实意义和作用，培养严谨认真的专业态度</p>	<p>1,2</p>	
<p>Logistic Regressi</p>	<p>经典工具的分类和回</p>	<p>2</p>	<p>案例讲解和讨论</p>	<p>应用于自</p>	<p>认识到 LogisticRegr</p>	<p>1,2</p>	

on 的分 类和回 归	归应用		论	选文本	ession的双重 现实意义和作 用, 培养严谨 认真的专业态 度	
技术 的 特 殊 应 用	结构化预测和异常检测	4	案例讲解和讨论	应用于自 选文本并 比较	掌握技术特殊 应用的语言学 意义, 认知专 业知识的社会 意义和价值	3,4,5,6
聚类	无监督学习方法的应 用	2	案例讲解和讨论	应用于自 选文本并 比较	掌握聚类方法 的语言学意 义, 认知专业 知识的社会意 义和价值	3,4,5,6
其 他 机 器 学 习 模 式	系统学习机器学习方 法的补充	2	案例讲解和讨论	应用于自 选文本并 比较	掌握更多机器 学习方法的语 言学意义, 认 知专业知识的 社会意义和价 值	3,4,5,6
深 度 学 习	线性回归、文本分类、 logistic 回归等深度学 习方法	2	案例讲解和讨论	应用于自 选文本并 比较	掌握深度学习 方法的语言学 意义, 认知专 业知识的社会 意义和价值	3,4,5,6
机 器 学 习 语 言 模 型	语言模型模式的跨语 言模型、语料库微调、 模型组合等	2	案例讲解和讨论	应用于自 选文本并 比较	掌握机器学习 语言模型的语 言学意义, 认 知专业知识的 社会意义和价 值	3,4,5,6
案 例 《 红 楼 梦 》 前 80	通过机器学习方法识 别文本异同	2	案例讲解和讨论	应用于自 选文本并 比较	掌握机器学习 与经典文本相 结合的意义,	3,4,5,6

	回 和 后 40 回					认知专业知识的 社会意义和 价值	
	讲 好 中 国 故 事	训练自己的模型	2	案例讲解和讨 论	应用于自 选文本并 比较	掌握语言模型 的语言学意义 和重要性, 认 知专业知识的 社会意义和价 值	3,4,5,6
	重读《共 产 党 宣 言》	重要社会概念的演变	2	案例讲解和讨 论	应用于自 选文本并 比较	掌握语言模型 的语言学意义 和重要性, 认 知专业知识的 社会意义和价 值	3,4,5,6
	语 言 学 变 量 及 其 协 同 效 应	认识并设置语言学 研究变量	2	案例讲解和讨 论	应用于自 选文本并 比较	掌握变量的语 言学研究意义 和重要性, 认 知专业知识的 社会意义和价 值	3,4,5,6
注 1: 建议按照教学周周学时编排。							
注 2: 相应章节的课程思政融入点根据实际情况填写。							
<b>*考核方式</b> (Grading)	示例: (1) 平时作业 30 分 (2) 期末大作业 70 分						
<b>*教材或参考资 料</b> (Textbooks & Other Materials)	自编教材						
其它 (More)							
备注 (Notes)							