

#### 目录

精通数据科学

### ONE 朴素贝叶斯模型

伯努利模型与多项式模型

精通数据科学。

从给收益回的秘证不管管

精通数据科学

精通数据科学

TVVO Pipeline 模型的联结

精通教师和潜。

THREE 中文分词。如果是是

更精确地特征提取

#### 朴素贝叶斯模型

模型数据

精通数据科学。从始级回归和深度证明

模型数据存储结构

C3-Art/

C3-Art0002.txt

C4-Art0003.txt

...

C4-Literature/

C4-Literature0001.txt

C4-Literature0004.txt

...

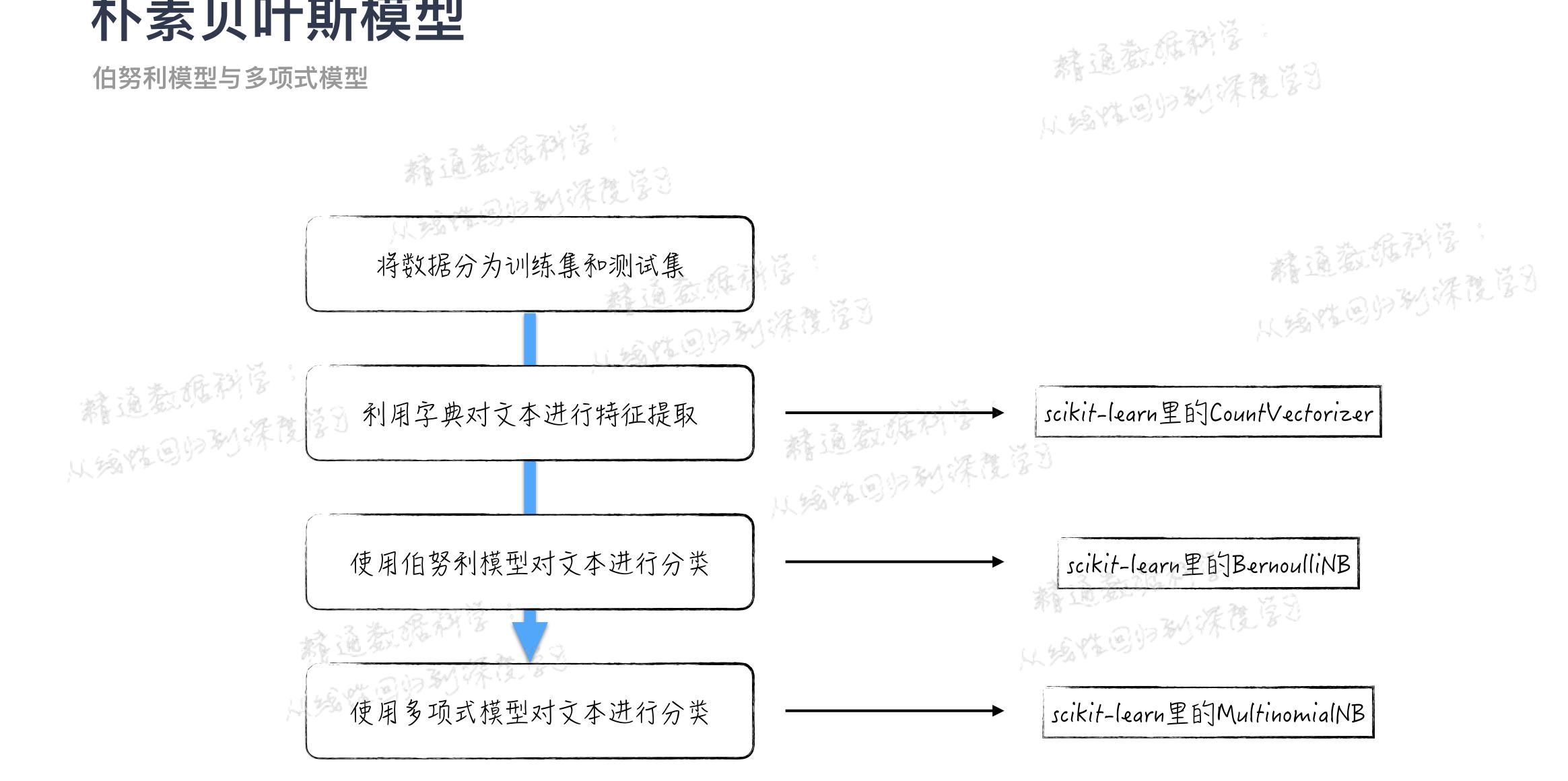
C5-Education/

模型数据来源于复旦大学计算机信息与技术系国际数据库中心自然语言处理小组(李荣陆提供)

3

#### 朴素贝叶斯模型

伯努利模型与多项式模型



#### 目录

精通激派和学 

伯努利模型与多项式模型。

精通数纸料管。 从给此回的秘珠度管

精通数据科学。

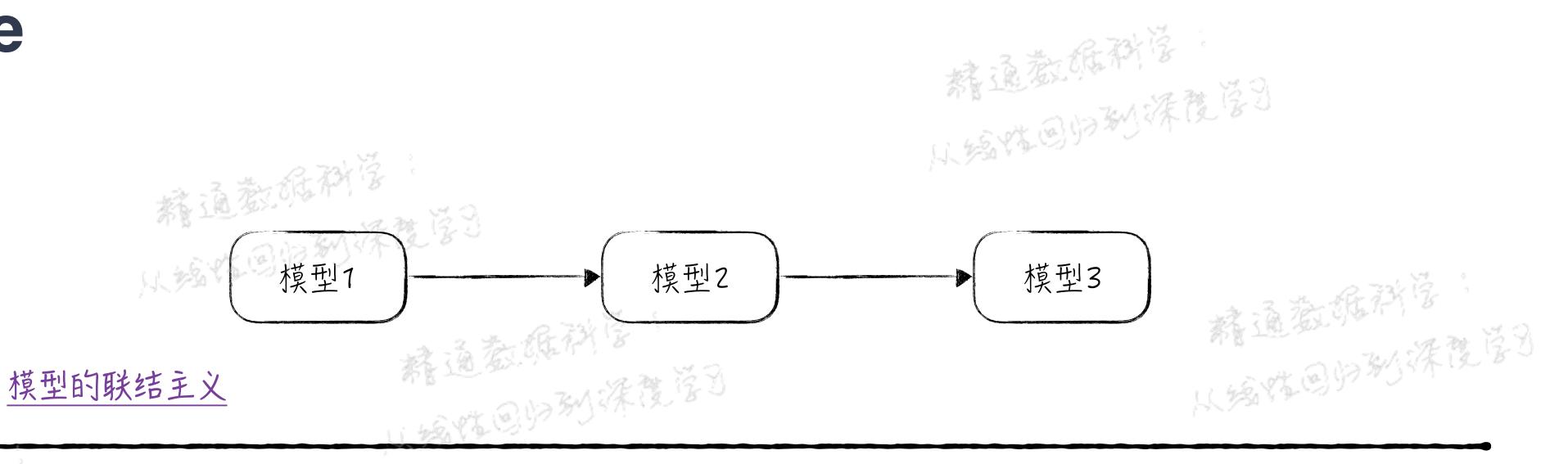
精通数据科学 从线性回归初深度管

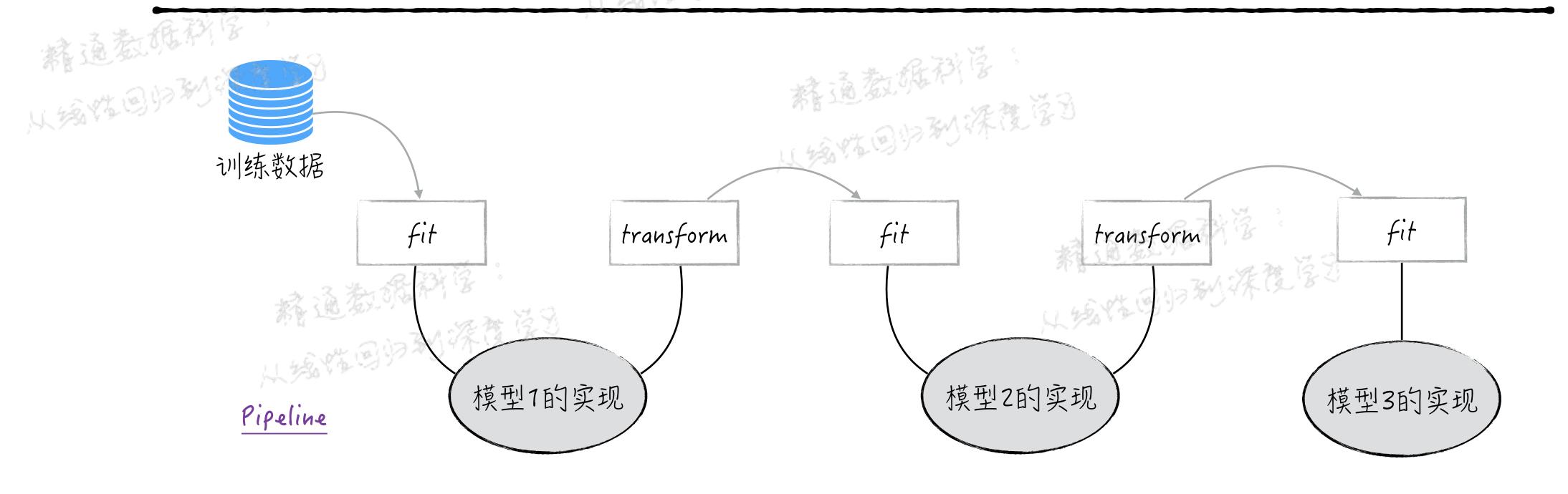
TVVO Pipeline 模型的联结

精通数据科学 从始级国的强烈深度 THREE 中文分詞の分科技

#### Pipeline

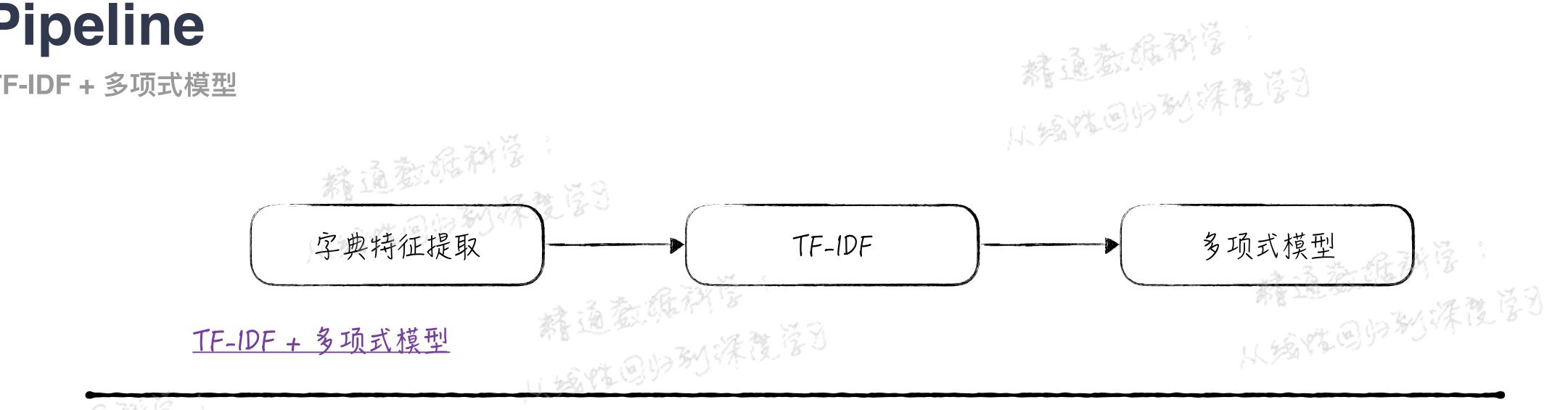
模型的联结

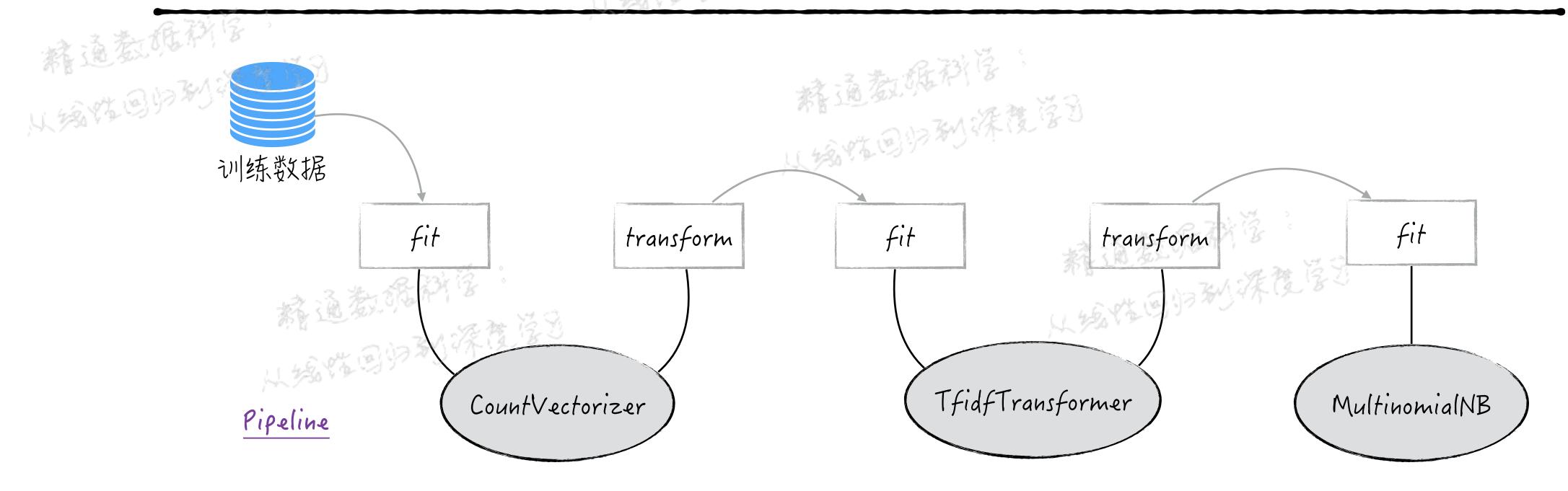




#### Pipeline

TF-IDF + 多项式模型





#### 目录

精通数据科学 

伯努利模型与多项式模型。

精通数纸料学。 从绝对国的秘证不改管的

> 精通数据科学 从绝对回的粉件度管的

精通数据科学 从线性回归初深度管的

TVVO Pipeline 模型的联络 (State of the August Au

精通数据科学: 从细胞回的多对深度管 THREE 中文分词。

更精确地特征提取

#### 中文分词

jieba分词

精通数据科学。

## 使用pip install jieba安装jieba分词

不分词 + 伯努利

	14				
	不分证	月+伯努利	糖通数据科学		
表表据科学	precision	recall	f1-score	support	
C11-Space	度管型 0.57	0.84	0.67	207	品物质
C19-Computer	0.89	0.88	0.89	389	C1
C3-Art	0.77	0.71	0.74	222	12 Pm
C39-Sports	0.79	0.62	0.69	381	14 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
micro avg	0.76	0.76	0.76	1199	
macro avg	0.75	0.76	0.75	1199	
weighted avg	0.78	3 20.76	0.76	1199	we

			查纸纸料管	0 0			
3	分词+伯努利						
	precision	recall	f1-score	support			
C11-Space	0.92	0.82	0.87	207			
C19-Computer	0.83	0.99	0.90	389			
C3-Art	0.97	0.72	0.83	222			
C39-Sports	0.86	0.86	0.86	383			
micro avg	0.87	0.87	0.87	1199			
macro avg	0.89	0.85	0.86	1199			
	Action to the second se	0.87	0.87	1199			

更好地特征提取,模型效果提升明显

特征提取是建模成 功的关键

精通数据科学。 从验验证到的秘证不改资

# THANKSOUS

務通数据科学 从给您回归和深度管

村通教师和强。

精通数据科学。 从绝路的多处深度管

精通数据科学

精通数派科学 从给你回的秘况