
泰迪科技数据挖掘项目实训

多种农产品价格智能预测

培训解决方案

广东泰迪智能科技股份有限公司 版权所有

地址：广州市经济技术开发区开泰大道 36 号 1 栋 212

网址：<http://www.tipdm.com>

邮箱：services@tipdm.com

邮编：510663

联系人：

电话：

目录

1 项目介绍.....	3
1.1 项目背景.....	3
1.2 项目目标.....	3
1.3 项目数据.....	4
1.4 项目周期及时间安排.....	5
1.5 项目难度.....	5
2 项目任务.....	6
3 项目流程.....	6
4 项目核心.....	6
5 实现工具.....	6
6 实训对象.....	6
7 前置知识.....	6
8 项目提交规范.....	7
8.1 项目交付内容要求.....	7
8.2 项目交付命名规范.....	7
8.3 项目提交邮件格式.....	8
9 实训对应的就业岗位.....	8
9.1 就业岗位.....	8
9.2 岗位分析.....	9
10 项目收获.....	9
11 项目评价及积分.....	9
12 附件一 工作室管理制度.....	11
13 附件二 前置课程课表.....	11

1 项目介绍

1.1 项目背景

中国自古以来就是一个农业大国，农业关系到国计民生，是国民经济的支柱产业。农产品价格是否合理，不仅影响农业生产的发展，农产品的流通、消费和农民的收入水平，而且影响工业品的成本和价格，影响国家同农民之间、城乡人民之间以及农民内部的物质利益关系，对整个社会经济生活的安定也关系重大。

与其他一般商品不同，农产品价格受到气候、供求关系、宏观经济等因素的影响后波动性较大。近年来，有关国内农产品涨价的信息频频见诸媒体，有些城市已经吃不到 3 元以下的蔬菜了，而春节前批发价只有五六元一斤的绿豆，如今接近 10 元钱。最疯狂的涨价产品是大蒜，借“甲流”概念疯涨过后，又卷土重来，据报道，河南大蒜身价短时间内涨了 100 倍。针对波动幅度如此之大的农产品价格，能否对其进行宏观调控与管理，不仅关系到广大农民的生计问题，更是影响国计民生的头等大事。

如果能够通过大量历史数据的分析，预测未来的价格走势，就能为决策者提供更有力的数据支持，从而制定出更为科学合理的政策。

1、提交结果

提交所选取的 935 种农产品所预测的价格 excel 文件（包含农产品名称）。

2、主流技术

本项目主要涉及技术为利用 Pandas 库进行数据预处理，缺少值填充，利用 sklearn 模块建模模型并预测。

1.2 项目目标

利用填补后的数据构建模型，对所选取出的 935 种农作物价格做出预测。

1.3 项目数据

本案例数据来源为 2016 年以前全国各农场品交易市场，数据中保留了交易市场所在地，市场名称，农场品类别、名称、最低交易价格、评价交易价格、最高交易价格、抓取时间、交易时间。在提供的数据中，已经对市场名称、农产品名称做了映射处理。数据样例如图 1-1 所示。数据提供方除了提供 2016 年以前全国各农场品交易市场农产品交易数据外还给出了 935 种农产品及其对应的交易市场，并对需要预测的日期做出要求。预测目标数据如图 1-2 所示。

农产品市场名称	农产品名称	农产品规格	区域	颜色	单位	最低交易价	平均交易价	最高交易价	数据入库时间	数据发布时间
云南	F84FFE61玫瑰	9FABF209C\N	\N	红色	20支/扎	19	22	24	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	5E0CB63A2\N	\N	红色	20支/扎	18	22	24	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	FC7FE9E4A\N	\N	粉色	20支/扎	16	19	22	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	C6D05CB7C\N	\N	黑红色	20支/扎	16	18	20	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	427BAA73E\N	\N	粉色	20支/扎	20	24	26	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	592513E0A\N	\N	白色	20支/扎	15	18	19	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	7771972EA\N	\N	白色	20支/扎	10	20	25	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	0E5854EEF\N	\N	香槟色	20支/扎	18	24	26	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	A6EE72B6E\N	\N	桃红色	20支/扎	18	20	25	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	19EAE737F\N	\N	深紫色	20支/扎	16	20	24	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	EC23A1984\N	\N	紫色	20支/扎	10	15	18	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	711C97C4E\N	\N	蓝色	20支/盒	27	28	30	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	6DB8A93D\N	\N	各色	20支/盒	27	28	30	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	B44C311E\N	\N	粉色	20支/扎	20	26	28	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	AD081A45F\N	\N	橙色	20支/扎	20	26	30	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	E7DA9559C\N	\N	黄色	20支/扎	12	18	20	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	B412A5B5E\N	\N	花边色	20支/扎	10	15	18	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	8B542827E\N	\N	杏红色	20支/扎	12	20	25	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	3815F5EFC\N	\N	乳白色	20支/扎	15	20	25	2016/7/18 15:05	2015/11/27
云南	F84FFE61玫瑰	70FAC8E7E\N	\N	深红色	20支/扎	15	18	22	2016/7/18 15:05	2015/11/27

图 1-1 数据样例

农产品市场名称	农产品名称	农产品规格	区域	颜色	单位	数据入库时间	数据发布时间
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/1	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/2	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/3	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/4	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/6	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/7	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/8	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/9	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/10	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/11	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/12	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/13	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/14	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/15	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/16	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/17	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/18	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/20	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/21	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/22	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/24	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/25	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/26	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/27	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/28	
广西	31D831D03A02090839BF1076E31DC00E	粮油 饲料	OC63912C7F6775380A4AB1D78E8A3FD3	null	null	34:18. 2016/1/29	
江苏	5261587C6B5E768AACFF8EB6B920CAD3	粮油 饲料	D148E47ADD84F353C7D7BE4EA4528FF1	null	null	34:18. 2016/1/1	
江苏	5261587C6B5E768AACFF8EB6B920CAD3	粮油 饲料	D148E47ADD84F353C7D7BE4EA4528FF1	null	null	34:18. 2016/1/2	
江苏	5261587C6B5E768AACFF8EB6B920CAD3	粮油 饲料	D148E47ADD84F353C7D7BE4EA4528FF1	null	null	34:18. 2016/1/3	
江苏	5261587C6B5E768AACFF8EB6B920CAD3	粮油 饲料	D148E47ADD84F353C7D7BE4EA4528FF1	null	null	34:18. 2016/1/4	

图 1-2 预测目标数据

1.4 项目周期及时间安排

注：以下周期仅为参考，工作室可在选择该项目后自行选择项目开始和结束时间，尽量控制在规定周期内完成，以确保学生可以达到训练目的。

(1) 项目周期：2 周

(2) 时间安排

任务安排	内容	提交场所	提交时间
第一周	任务 1 明确项目需求与目标	邮箱	Day1 22:00
	任务 2 环境准备：numpy、pandas、sklearn（5%）		
	任务 3 数据探索（作业代码）（20%）		Day5 22:00
	任务 4 数据预处理（作业代码）（40%）		Day1 22:00
	任务 5.1 样条插值法理解（原理阐述）（50%）		Day5 22:00
第二周	任务 5.2 方法实现（作业代码）（60%）		
	任务 6 模型选择与建模（作业代码）（70%）		Day1 22:00
	任务 7 模型调参与评价（作业代码）（90%）		
	任务 8 完成项目报告（作业）（100%）		Day5 22:00
在线答疑		QQ 群	

1.5 项目难度

★★★

2 项目任务

任务 1 明确项目需求与目标

任务 2 环境准备：numpy、pandas、sklearn 等

任务 3 数据探索（作业代码）

任务 4 数据预处理（作业代码）

任务 5 缺失值补全（作业代码）

任务 5.1 样条插值法理解

任务 5.2 方法实现

任务 6 模型选择与构建

任务 7 模型调参与评价（作业代码）

任务 8 完成项目报告（作业）

3 项目流程

- (1) 明确项目目标：阅读项目任务书，明确项目交付内容。
- (2) 学习前置知识：提供项目所需知识的云课堂课程，快速掌握项目前置知识。
- (3) 项目实践：获取项目需求，动手做项目。
- (4) 项目验收：对项目成果进行验收，指导改进项目成果。
- (5) 信息入库：参训学生信息录入泰迪人才库，为优秀结业生引荐合作公司。

4 项目核心

- Python 数据分析
- Python 机器学习

5 实现工具

numpy、pandas、sklearn 等

6 实训对象

数学、统计学、计算机等相关专业学生。

7 前置知识

第 1 模块：Python 编程基础

第 2 模块：Python 数据分析与应用

第 3 模块：Python 机器学习实战

8 项目提交规范

8.1 项目交付内容要求

(1) 请在“项目名称(模板).docx”基础上进行报告的编写。

(2) 请在“项目名称(模板).pptx”基础上进行 PPT 的编写。

(3) 具体的内容以及格式要求请查看课程【工作室项目交付要求讲解】。

注：请按照课程【工作室项目交付要求讲解】中提及的要求进行文档的编写，若不按要求进行规范编写，提交的项目文件将被返回修改。

8.2 项目交付命名规范

(1) 项目交付文件包括中间数据（如有）、代码、报告、PPT。

(2) 中间数据需放入“01.中间数据（如有）”文件夹中，各中间数据文件命名为“任务 x-x”，如“任务 1-1”。

(3) 代码需放入“02.代码”文件夹中，各代码文件命名为“任务几 任务名称”，如“任务 1 预处理航空客户数据”。

(4) 报告需放入“03.报告”文件夹中，报告命名为“项目名称”，如“航空公司客户价值分析”。

(5) 报告需放入“04.PPT”文件夹中，PPT 命名为“项目名称”，如“航空公司客户价值分析”。

(6) 最终把所有文件夹放入总文件夹中，总文件夹命名为“项目名称--院校工作室名称--小组成员名称（用顿号隔开）--上交日期（如 2021.1.23）”，如“航空公司客户价值分

析--重理工智能工作室--刘备、关羽、张飞--2021.1.23”。

(7) 最后仅需把(6)提及的总文件夹的压缩包提交至邮箱 taididingyuehao@tipdm.com。

注：具体也可查看课程【工作室项目交付要求讲解】中的项目交付模板。同时，请严格按照以上方式进行上缴，如上缴方式不正确，可能影响评级。

8.3 项目提交邮件格式

- (1) 邮件收件人为：taididingyuehao@tipdm.com
- (2) 邮件主题名称为：院校工作室名称
- (3) 邮件正文内容为：项目名称-小组成员名称
- (4) 邮件附件为：“8.2 项目交付命名规范”提及的项目总文件夹压缩包

项目提交邮件格式具体参考如下。

收件人	<input type="text" value="taididingyuehao@tipdm.com;"/>
添加抄送 - 添加密送 分别发送	
主题	<input type="text" value="T0001泰迪·黔南师院数统智能工作室"/>
继续添加 插入正文 表情 正文模版 截屏 样式1	
正文	<div><p>B <i>I</i> <u>U</u> 链接 插入 列表 引用 代码 更多 HTML</p><p>你好：</p><p> 本周完成</p><p> 1.Python供应链商品销售数据分析-成员为张三、李四</p><p> 2.Python供应链商品销售数据分析-成员为小青、小白</p><p> 3.Python新冠疫情数据分析-成员为小红、小明</p><p> 详见附件，请查收</p><p>此致！</p></div>

9 实训对应的就业岗位

9.1 就业岗位

数据分析师、数据挖掘工程师。

9.2 岗位分析

序号	岗位	主要业务工作	所需技能	相应课程设置
1	数据分析师	数据处理、分析建模、撰写分析报告	Python 数据分析 分析报告	Python 语言设计 Python 数据分析与应用 数据库
2	数据挖掘工程师	模型开发、数据分析处理	Python 数据分析 挖掘建模	Python 语言设计 Python 数据分析与应用 数据库入门 机器学习

10 项目收获

- 足不出校门即可获得实战技能。
- 了解数据分析岗位目前的就业形式和前景，了解需要掌握的技能。
- 掌握一定的挖掘技能和工具，体验一个实际项目的全过程。
- 参训学生信息录入泰迪人才库，为优秀结业生引荐合作公司。

11 项目评价及积分

(1) 公司将按照工作室对该项目的完成情况给予评分，项目评分由低到高依次为 D、C、B、B+、A、A+，B 为及格，B 以下不及格。

(2) 企业将根据工作室学员的项目评分给予工作室对应的积分，积分达到一定水平，可以获取更多泰迪资源。

(3) 可获得的资源包括但不限于

A.由泰迪科技颁发的 CBDA 证书



B.由泰迪科技颁发的实习证明



C.云课堂课程资源

(4) 评分对应积分如下

项目评分	积分
项目评分为 B	+50
项目评分为 B+	+60
项目评分为 A	+70
项目评分为 A+	+80

(5) 工作室制度详见附件一。

12 附件一 工作室管理制度

查看网页: <https://kdocs.cn/l/ckdaaSEuhp0y>

13 附件二 前置课程课表

Python 编程基础	Python 数据分析与应用
1 准备工作	1 Python 数据分析概述
1.1 认识 Python	1.1 认识数据分析
1.2 搭建 Python 环境	1.2 熟悉 Python 数据分析的工具
1.3 安装 PyCharm	1.3 安装 Python3 的 Anaconda 发行版
1.4 PyCharm 使用入门	1.4 掌握 Jupyter Notebook 常用功能
2 列表操作	2 NumPy 数值计算基础
2.1 第一个 Python 程序	2.1 认识 NumPy 数组对象 ndarray
2.2 Python 固定数据类型介绍	2.2 认识 NumPy 矩阵与通用函数
2.3 列表构建及索引操作	2.3 利用 NumPy 进行统计分析
2.4 列表元素的增删改查操作	3 Matplotlib 数据可视化基础
2.5 列表推导式	3.1 了解绘图基础语法与常用参数
3 程序流程控制语句	3.2 分析特征间的关系
3.1 Python 常用操作符	3.3 分析特征内部数据分布与分散状况
3.2 Python 条件判定语句	4 Pandas 统计分析基础
4 字符串操作	4.1 读写不同数据源的数据
4.1 字符串及其索引&切片	4.2 掌握 DataFrame 的常用操作
4.2 字符串的常见方法	4.3 转换与处理时间序列数据
4.3 字典的创建及索引	4.4 使用分组聚合进行组内计算
4.4 字典常用操作	4.5 创建透视表与交叉表
4.5 字典推导式	5 使用 Pandas 进行数据预处理
5 Python 文件读取操作	5.1 合并数据
5.1 Python 读取文件	5.2 清洗数据
6 函数	5.3 标准化数据
6.1 Python 函数自定义	5.4 转换数据
7 面向对象与模块	6 使用 scikit-learn 构建模型
7.1 Python 方法与函数对比介绍	6.1 使用 sklearn 转换器处理数据
7.2 Python 面向对象示例	6.2 构建并评价聚类模型

7.3 Python 模块使用 7.4 第三方库的安装与调用 8 注意事项 8.1 Python 工作路径说明 8.2 模块命名及存放路径的注意事项 8.3 结语	6.3 构建并评价分类模型 6.4 构建并评价回归模型
--	--------------------------------

Python 机器学习实战	
1 机器学习绪论 1.1 引言 1.2 基本术语 1.3 假设空间 1.4 归纳偏好 2 模型评估与选择 2.1 经验误差与过拟合 2.2 评估方法 2.3 性能度量 2.4 偏差与方差 3 回归分析(Regression Analysis) 3.1 基本形式 3.2 线性模型 3.3 逻辑回归 4 决策树(Decision Tree) 4.1 基本流程 4.2 划分选择 4.3 剪枝 5 神经网络(Artificial Neural Network) 5.1 神经元模型 5.2 感知机与多层网络 5.3 误差逆传播 5.4 BP 神经网络	

6 最近邻算法 (KNN) 7 朴素贝叶斯分类(Naive Bayesian) 8 聚类分析(Cluster Analysis) 8.1 聚类任务 8.2 性能度量 8.3 距离计算 8.4 常用聚类算法 9 支持向量机(Support Vector Machine) 9.1 间隔与支持向量 9.2 对偶问题 9.3 核函数 9.4 软间隔与正则化	
--	--

工作室邀请函

为了适应大数据与人工智能及发展的需求，顺应教育部提倡的深化校企合作的号召，更好的服务于广大数据分析爱好者，广东泰迪智能科技股份有限公司诚邀各高校相关专业老师、相关协会学会、俱乐部等组织合作成立“泰迪·智能工作室”，工作室以独立的模式运营，并以学生为中心成立，受泰迪科技监督且由其免费提供各种工作室所需资源的创新型数据智能工作室。



工作室旨在通过教育与产业之间的联动，实行“引进来，走出去”模式，引导学生学习数据科学与人工智能方法为导向，通过与企业的联系、合作、实践，激发学生的数据分析思维，全面推进数据分析与人工智能发展，提高大学生的数据分析素质，激发学生的创新创业精神，以实现创新型数据智能创业人才为培养目标。