

|  |  |
| --- | --- |
| 题 目： | 基于深度学习的猫品种的消费行为及其分类研究 |
| 学 院： | 数据科学学院 |
| 专 业： | 数据科学与大数据技术 |
| 年级班别： | 2021级数据科学与大数据技术班 |
| 学 号： | 421470155 |
| 学生姓名： | 葸雯丹 |

**提交日期：** 2024 **年** 11 月

**目录**

[课程论文 1](#_Toc24082)

[绪论 3](#_Toc4327)

[1.1研究背景与意义 3](#_Toc5162)

[1.2 研究现状 4](#_Toc15499)

[1.3 研究目标 4](#_Toc23710)

[1.4 宠物经济概述 5](#_Toc25007)

[CNN分类算法 8](#_Toc20431)

[2.1对数据进行预处理 8](#_Toc17893)

[实现SVM分类器 8](#_Toc14577)

[Random Forest分类器 10](#_Toc23507)

[2.3数据处理及训练模型 10](#_Toc19277)

[Random Forest模型准确率为0.97 10](#_Toc20462)

[1. 准确率（Accuracy） 11](#_Toc30349)

[2. 精确率（Precision） 11](#_Toc7198)

[3. 召回率（Recall） 11](#_Toc31314)

[4. F1分数（F1Score） 11](#_Toc24939)

[1. 整体性能: 12](#_Toc7088)

[2. 精确率和召回率: 12](#_Toc33)

[3. 模型的适用性: 12](#_Toc13117)

[4. 选择模型时考虑实际需求: 12](#_Toc27153)

# 绪论

猫作为人类宠物的历史悠久，可以追溯到数千年前。自从人类文明出现以来，猫一直扮演着重要的角色，不仅作为家庭伴侣，还在农耕社会中帮助控制鼠灾。

1.1研究背景与意义

随着社会经济的不断发展和人们生活水平的提高，宠物经济逐渐成为消费市场的一个重要组成部分，其中宠物猫市场表现尤为突出。据统计，近年来，全球宠物猫市场的规模持续扩大，宠物猫数量和养猫家庭的比例不断上升。猫作为宠物不仅仅提供陪伴，还对人们的心理健康产生积极影响。因此，研究宠物猫市场的消费行为对理解消费者需求、优化市场策略具有重要意义。

宠物经济的重要性体现在多个方面。首先，宠物经济涵盖了宠物食品、医疗、用品、服务等多个产业链环节，直接推动了相关行业的发展。其次，宠物经济对社会文化产生深远影响，宠物文化逐渐受到重视，养宠行为成为一种新的生活方式和社会现象。再者，宠物经济在一定程度上反映了社会的经济水平和文明程度，是社会进步的一个标志。

在宠物猫市场中，不同猫品种的消费行为差异研究尤为重要。不同猫品种具有不同的外貌特征、性格特点和生活习性，这些差异会在一定程度上影响消费者的购买决策和消费行为。例如，有的消费者可能更倾向于购买外形优雅的暹罗猫，而有的消费者则可能更喜欢性格温顺的布偶猫。研究不同猫品种间的消费行为差异，有助于深入理解消费者的心理需求和偏好，从而更好地为市场提供个性化的产品和服务。

1.2 研究现状

目前，对于宠物消费行为的研究已初具规模，大量学者从不同视角进行了相关研究。

宠物消费行为的基本理论研究。许多研究基于消费者行为理论，从动机、需求、决策过程等角度分析了宠物消费行为的基本规律。例如，有研究指出，养宠动机不仅包括情感陪伴需求，还包括社会身份象征等复杂因素。

市场细分研究。已有研究通过市场调研，尝试对宠物市场进行细分，识别出不同类型的宠物消费者及其特征。这些研究通过问卷、访谈等方法，获取大量消费者数据，进而分析消费者在宠物选择上的偏好、影响因素等。尽管已有研究为宠物消费行为提供了丰富的理论和实践基础，但针对不同猫品种之间的消费行为差异研究相对较少。现有研究大多集中在宠物犬的消费行为上，对宠物猫尤其是不同品种的细分市场缺乏深入分析。此外，很多研究未能结合现代数据分析技术，对大规模数据进行系统化分析，研究结论的普适性和精确性有待提升。

1.3 研究目标

1.确认不同猫品种对消费者购买决策的影响因素。这包括不同猫品种的外貌、性格、生活习性等特征对消费者选择的影响。

2.了解消费者在选择不同猫品种时的具体消费行为模式。例如，消费者在选择不同品种猫时的品牌偏好、价格敏感度、购买决策过程等。

3.分析消费者面对不同猫品种时的心理需求和偏好，识别出影响消费者购买决策的关键心理因素。

4.提出基于研究结论的市场策略建议，为相关企业提供决策参考，以优化宠物猫市场的产品和服务。

**1.4 宠物经济概述**

宠物经济是衡量一个社会文化水平和经济繁荣程度的一个重要标志。随着人们生活水平的提高，宠物不仅仅是人们的陪伴，更成为了一种家庭成员的象征。近年来，宠物市场的规模不断扩大，其中以宠物猫和宠物狗为代表的宠物市场发展尤为迅速。根据市场调研，全球宠物市场规模在近十年内保持了稳定的增长，年均增长率约为5%8%。其中，中国的宠物市场发展尤为迅猛，近年来更是展现出强劲的增长势头。

自2000年以来，中国的宠物猫数量显著增加。根据多个来源的报道，中国的宠物行业自2010年以来经历了爆炸性的增长。数据显示，从2010年到2016年，中国宠物行业的复合增长率达到49.1%，根据2021年的数据显示，中国的猫和狗的数量从2017年的8750万增长到2021年的1.124亿。这期间，宠物猫的增长也伴随着整个宠物行业的扩展。例如，2015年至2022年，中国的宠物消费从725亿元人民币增长到3960亿元人民币。

下图1为2000年~2023年宠物猫增长率趋势

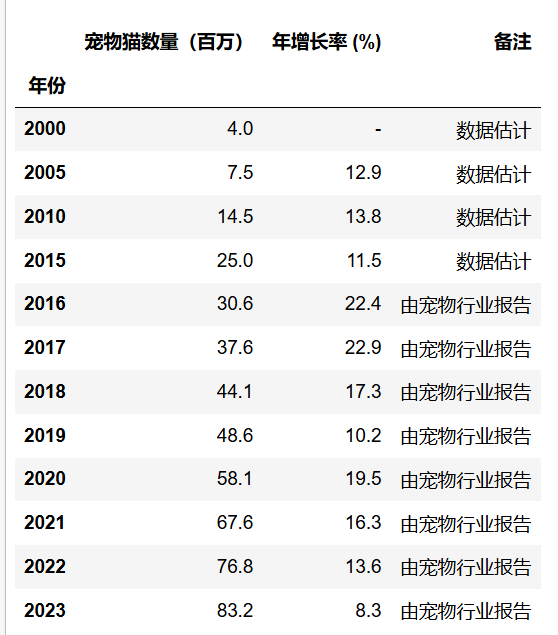


图1

下图1-1与1-2表明从2000年以来的宠物猫数量变化

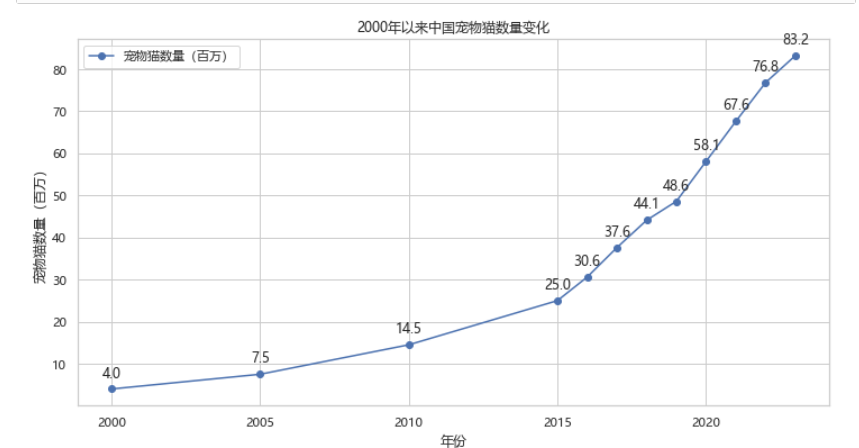


图1-1

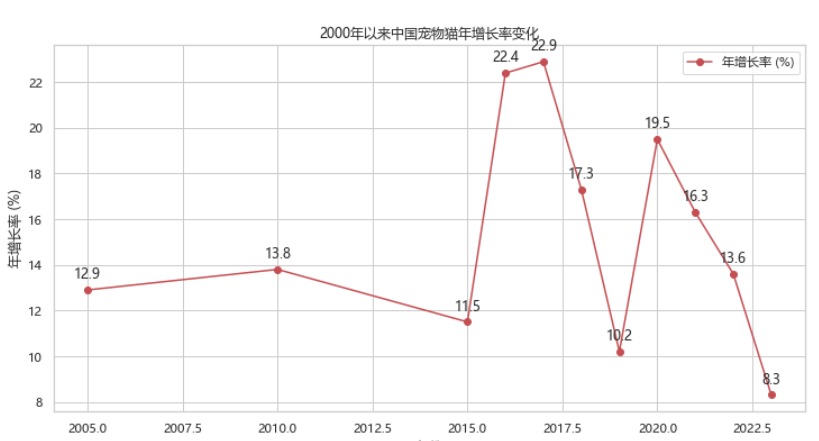


图1-2

CNN分类算法

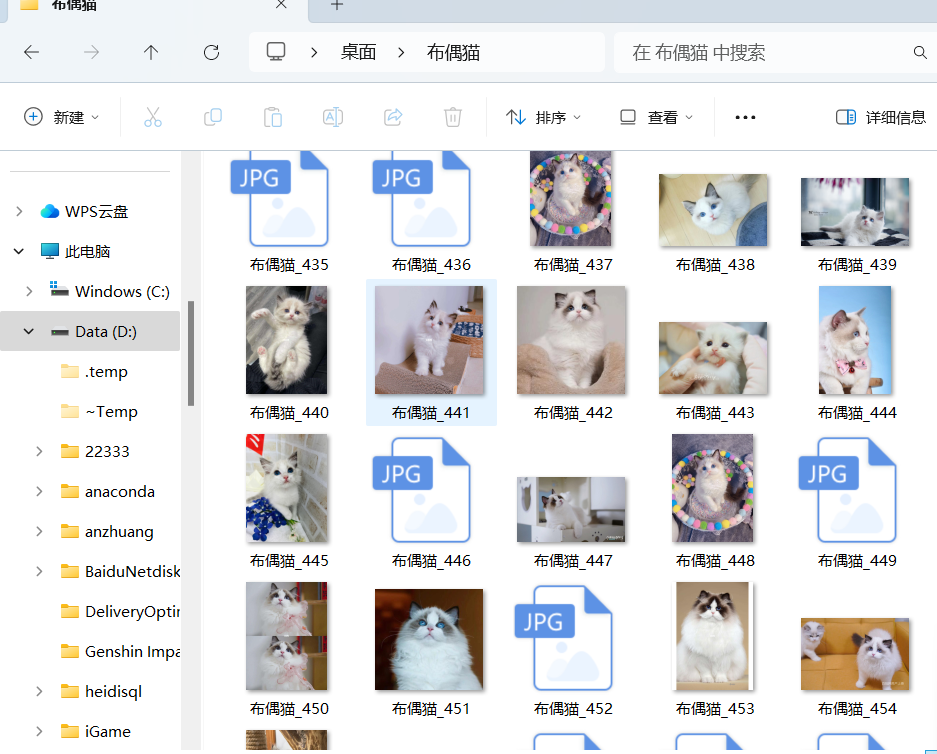
2.1图2-1表示对1300张图片进行预处理

图2-1

图2-2表示数据在CNN算法模型处理下的准确率

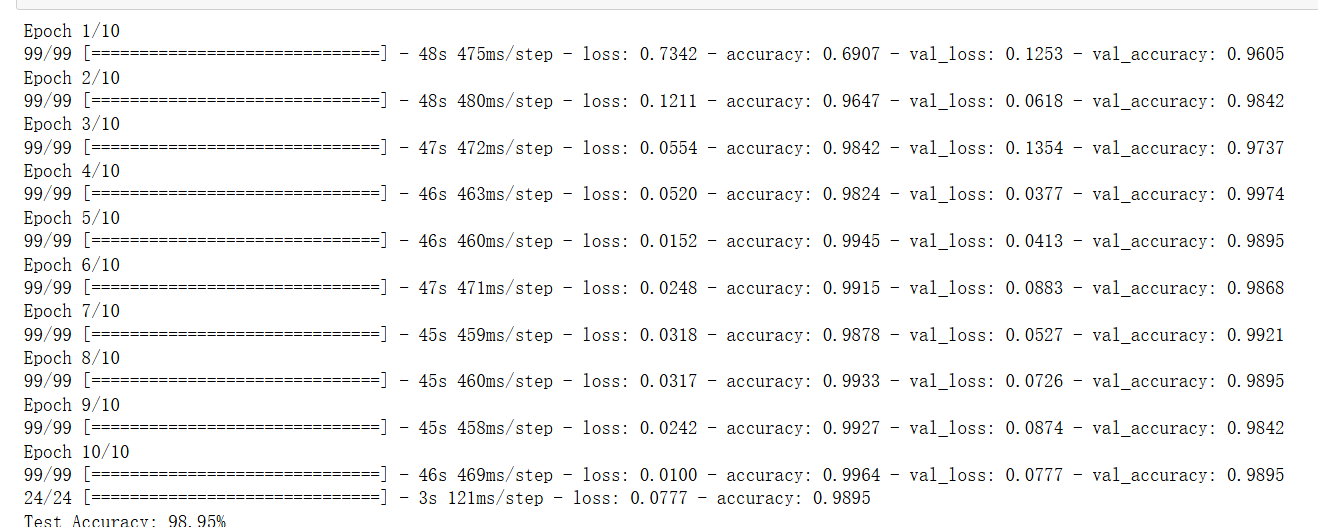


图2-2

CNN模型准确率为0.99

实现SVM分类器

2.2：对1300张图片进行预处理及训练模型

图3-1与图3-2表示数据SVM算法处理下的准确率



图3-1

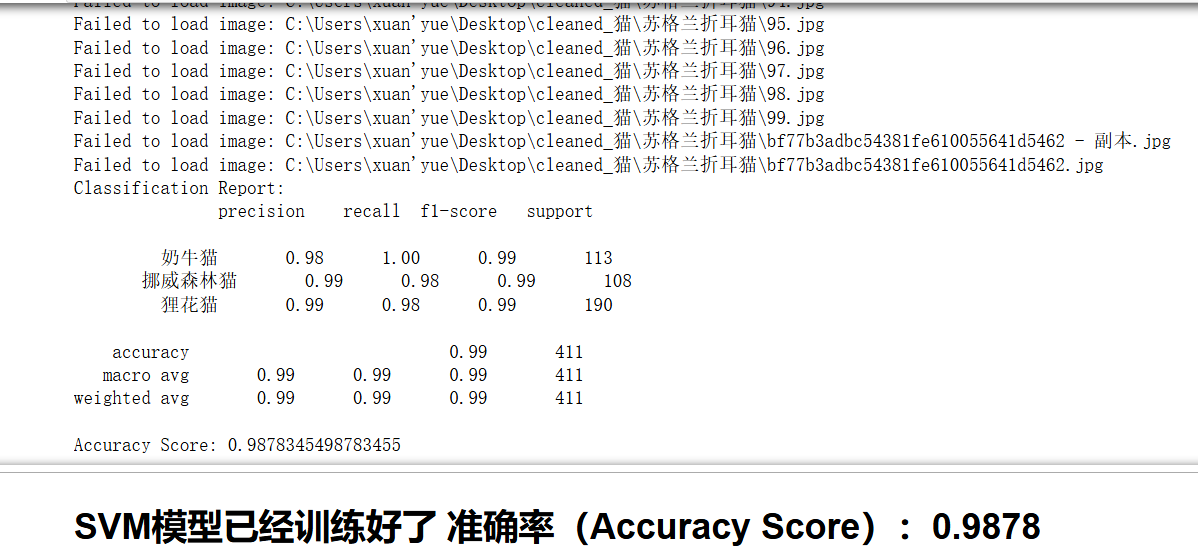


图3-2

Random Forest分类器

2.3对1300张图片进行处理及训练模型

下图4-1表示数据在Random Forest算法模型下的准确率

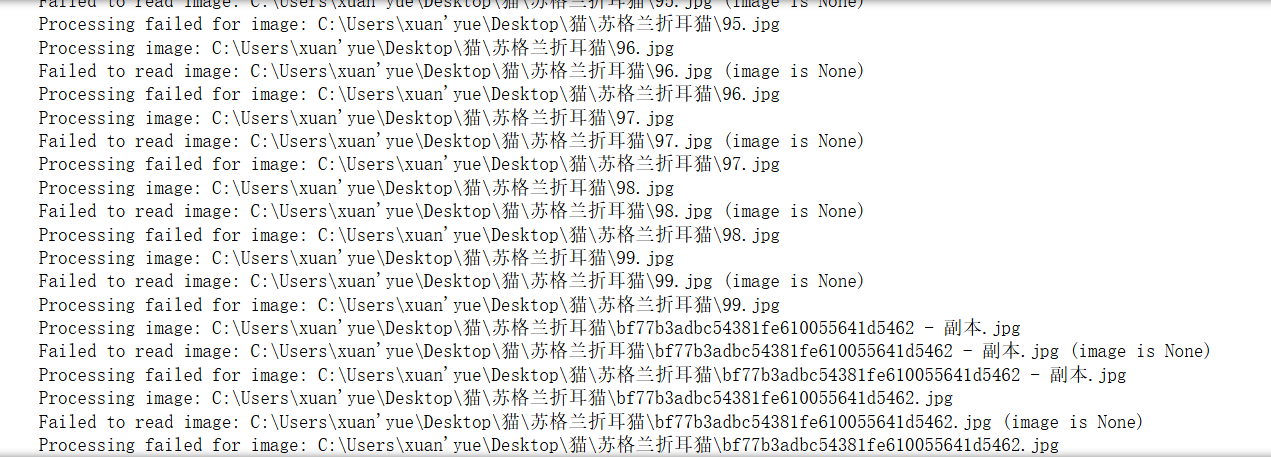


图4-1

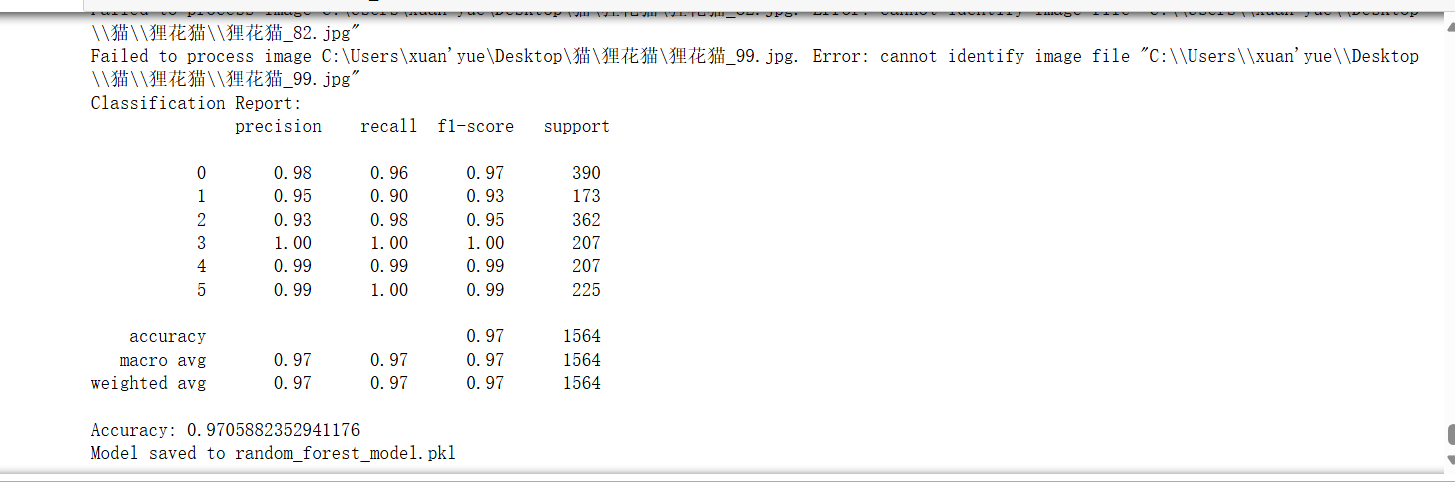


图4-2

Random Forest模型准确率为0.97

三：模型比较与建议

为了比较CNN、SVM和随机森林三种分类器的分类结果，使用以下分类指标进行分析：准确率（Accuracy）、精确率（Precision）、召回率（Recall）和F1分数（F1Score）。以下是具体比较结果：

1. 准确率（Accuracy）

CNN: 98.95%

SVM: 98.78%

随机森林: 97.06%

2. 精确率（Precision）

CNN: 平均精确率:0.98

SVM: 平均精确率: 0.99

随机森林:平均精确率: 0.97

3. 召回率（Recall）

CNN: 平均召回率:0.98

SVM：平均召回率: 0.99

随机森林:平均召回率: 0.97

4. F1分数（F1Score）

CNN: 平均F1分数: 0.99

SVM:平均F1分数: 0.99

随机森林:平均F1分数: 0.97

综合比较

CNN：在测试数据上的表现非常出色，准确率达到98.95%。 SVM：各项指标表现均衡，特别是奶牛猫、挪威森林猫和狸花猫的分类效果非常好，精确率、召回率和F1分数均为0.99。

随机森林：虽然总体准确率略低于CNN和SVM，但在类别3和类别4上的表现非常出色，精确率、召回率和F1分数均达到1.00。

结论

综合来看，CNN和SVM在分类性能上表现优异，尤其是在高准确率和均衡的分类指标方面。如果对某些特定类别的分类效果要求极高，可以考虑使用随机森林，因为其在某些类别上的精确率和召回率表现完美。

四、机器学习分类器的结论与建议

1. 整体性能:

CNN: 以98.95%的高准确率和预估接近0.99的F1分数表现出色。适用于对分类准确性要求极高的应用。

SVM: 平均F1分数为0.99，显示出非常均衡的性能，特别是在奶牛猫、挪威森林猫和狸花猫分类中表现尤为优异。

随机森林: 尽管整体准确率和F1分数略低于CNN和SVM，平均F1分数为0.97，但在某些类别上的表现达到1.00的完美水平。

2. 精确率和召回率:

CNN: 高精确率和召回率，适合处理复杂数据和图像分类任务。

SVM: 在精确率和召回率方面保持了很好的平衡，适用于分类较为均衡的场景。

随机森林: 在处理多类别分类任务时表现良好，尤其是在某些特定类别上精确率和召回率极高，适合在特定类别上要求更高的应用。

3. 模型的适用性:

CNN: 由于其强大的特征提取能力，适合复杂图像处理和需要自动化特征学习的场景，如计算机视觉任务。

SVM: 在高维度数据和中小型数据集上表现出色，适合特征维度较高但数据量适中的分类任务。

随机森林: 对于数据噪声具有较强的鲁棒性，适合处理数据多样性高且有不平衡类别的情况。

4. 选择模型时考虑实际需求:

对于高精度和复杂特征提取需求的任务，CNN是首选，特别是在图像分类领域。

如果数据维度高且需要稳定的分类性能，SVM是一个很好的选择，适用于文本分类和小规模数据集。

随机森林适用于数据多样且噪声较大的场景，尤其是需要处理多类别不平衡数据的任务。

针对特定类别进行定制化处理，以提高在这些类别上的精度和召回率。

CNN和SVM在精度和F1分数上的表现非常出色，适合大多数高要求的分类任务。

随机森林虽然在平均F1分数上略低，但其在某些特定类别上的表现非常优异，是处理多样性高和噪声数据的可靠选择。

在实际应用中，建议根据具体的应用场景和数据特征选择最适合的模型，并通过调优和评估来达到最优效果