****

盛安德敏捷软件工程训练营

项目结题报告

基于人工智能的风电功率预测软件开发

|  |  |  |
| --- | --- | --- |
| 班级 | 学号 | 姓名 |
| 电气2102班 | U202112266 | 曲嘉骏 |
| 电气2102班 | U202112261 | 李俊昊 |
| 电气2102班  电气2102班 | U202112254  U2021122XX | 董启翰  XXX |

电气与电子工程学院

2023年08月

# 1 研究背景和目标

## 1.1 国内外发展状况

国外风电功率预测技术起步较早，20世纪80年代就开始了风电功率预测相关技术研究，丹麦、西班牙、德国、美国等国家的风电预测技术已走在世界前列 ， 其研究出的风电预测系统（如丹麦开发整合的Zephry系统；德国奥尔登堡大学开发的Previento系统；德国太阳能研究所开发的风电功率管理系统(WPMS)；美国AWS Truewind公司开发的eWind风功率预测系统，等等）也相对成熟且均已投入运行。相对而言，我国对于风力发电系统功率预测的研究起步较晚，但是发展速度较快，在国家电网公司国家电力调度通信中心的组织下，中国电科院和吉林省电力公司开展了风电功率预测系统的研究和开发，并于2008年投入运行。目前已经开发出了基于人工神经网络、支持向量机等统计方法的风电功率预测模型，以及基于线性化和计算流体力学的物理模型，同时正在开展多种统计方法联合应用研究机统计方法与物理方法混合预测模型的研究。但是目前来看，全球各国对此的研究也存在部分缺陷，例如：测精度不满足电网调度运行的需要，预测的时间尺度不满足要求，风电场还没有建立风电功率预测系统，历史数据较为匮乏，以及风电功率预测系统成本较高等。

## 1.2 本课程的设计目标

本项目基于人工智能方法，利用Python等软件的编程功能以及建模功能，构建神经网络。结合气象参数，选取合适的输入特征，建立风电厂发电功率预测模型，大幅减少人工工作量及计算量的同时，实现对风电场发电功率的准确预测，使发电厂发电功率效率最大化。

我们希望开发一款风电功率可视化的软件，帮助用户更好理解风力发电特性，并帮助专业人士直观了解风电功率状况，实现短时风电功率预测。

# 2 团队任务完成情况

## 2.1 团队人员组成

"带我飞"小组是一个充满活力和创造力的团队，以曲嘉骏作为组长，李俊昊、董启翰和王乐天为组员。每个成员都拥有独特的技能和贡献，共同构建了一个协作紧密的团队。

## 2.2 任务分工

在"带我飞"小组中，每个成员都拥有明确的任务分工，以便充分发挥各自的专业技能和优势，共同推动项目的成功完成。

曲嘉骏负责数据处理和搭建深度学习模型：曲嘉骏将负责采集、清洗和预处理项目所需的数据，确保数据质量和可用性。他将运用他在深度学习领域的专业知识，设计并搭建适用于项目的深度学习模型，以实现对数据的准确预测和分析。

李俊昊负责软件UI设计和图形界面：李俊昊将专注于设计出直观、用户友好的软件界面，以确保用户能够轻松地使用项目提供的功能和服务。

董启翰负责模型测试与预测精度提升：董启翰将负责对深度学习模型进行全面的测试，发现潜在问题并提出改进方案。他还将致力于优化模型，提高其预测精度和性能。

王乐天负责收集数据与制作答辩内容：王乐天将负责收集项目所需的相关数据，并进行整理和管理。此外，他还将准备项目的答辩材料，包括制作演示文稿和准备口头演讲，确保在展示阶段能够清晰、有力地表达团队的成果和价值。

## 2.3 项目进度里程碑

1. 项目启动阶段：

确定项目开发目标、技术路线，并设计具体开发方案。

进行团队成员的分工，明确各自职责和角色。

2. 数据准备阶段：

收集风电数据集，包括风速、功率输出等相关数据。

对数据进行清洗、预处理和特征工程，以便用于模型训练。

3. 模型开发阶段：

选择适合的深度学习算法，LTSM、RNN、MLP、GRU、CNN。

开发模型的初始版本，并进行基本的训练。

4. 模型优化阶段：

进行模型的交叉验证和性能评估，识别潜在的过拟合或欠拟合问题。

调整模型参数，优化模型性能和预测精度。

5. UI界面设计与开发阶段：

设计用户界面，使其易于使用和导航。

开发图形界面，与模型集成，允许用户输入参数并查看预测结果。

6. 模型测试与验证阶段：

使用预留数据集对模型进行测试，评估其在真实数据上的表现。

确保模型在不同情况下的预测精度和稳定性。

7. 答辩准备阶段：

收集项目相关数据、结果和成果，准备答辩所需的材料。

准备演示文稿和演讲内容，以便能够清晰地展示项目成果。

8. 项目展示与答辩阶段：

向老师和同学们展示项目成果，包括预测结果和界面功能。

进行答辩，回答问题并演示项目的实际应用和潜在价值。

9. 项目总结与报告阶段：

撰写项目总结报告，记录项目的目标、方法、成果和经验教训。

总结团队在项目中取得的成功和面临的挑战。

10. 项目结束：

进行项目评估，收集团队成员的反馈，以便在未来改进项目管理和执行。

归档项目文档、代码和资料，确保相关信息得以保存和分享。

# 3 总体设计

## 3.1 多因素影响分析

在项目启动前我们分析和评估了多种影响因素，以确定最终软件可以达到的目标和可以实现的功能。

1. 数据的收集和处理：在数据收集的过程中，我们需要用专业知识确定输入数据的范围和种类；之后，我们需要确定数据处理的方法，通过准确的收集和处理提高数据质量；同时，也应收集风机的理论数据，以进行比对和后续的数据清洗。
2. 模型建立：我们需要查阅资料，找到几种在风功率预测领域表现较好的神经网络模型，并分别建立模型。
3. 误差分析：我们需要引入评估神经网络模型的几种主要指标，以检验模型的准确程度。
4. 预测功能：我们需要找到合适的预测方法，实现通过用户的输入数据预测短时风电功率的功能。
5. 界面开发：在可视化软件开发方面，我们需要找到合适的编程语言和开发工具，以设计可以实现预期功能并配备优良界面的软件。

## 3.2 多技术比较

在我们的软件开发中，我们需要选择合适的编程语言，前端界面开发工具及数据库，以在实现预期功能的同时提高工作效率。以下是我们对几项关键技术的比较和最终得出的方案：

1. 编程语言：我们在前期的数据处理部分使用了matlab，但考虑到在之后的算法开发中，python语言拥有更多且更为成熟的机器学习和深度学习库，我们便确定使用python作为唯一使用的编程开发语言，并将前期的matlab代码转化成python代码。
2. 界面开发工具：相比于独立还需另行下载的Qt creator，我们选择了python的PyQt5中自带的Qt Designer作为软件界面的开发工具。我们在可视化的GUI上设计界面，而且可以直接自动生成对应窗口的py文件，交互方便，大大提升了工作效率。
3. 数据格式：由于数据是一次性存取，相比于调用和查询缓慢的数据库，我们决定使用csv格式存储数据，以便导入程序和之后的开发测试。

## 3.3 总体设计方案

基于上述分析，我们最终确定了总体设计方案如下：

1. 数据收集和处理：我们确定收集风机的瞬时风速，30s平均风速，10min平均滑动风速，风向，温度，桨距角和有功功率数据，并通过以下规则对数据进行处理：

（1）筛除风机运行状态异常的数据点

（2）筛除切入风速以下，切出风速以上且功率不为0的异常数据

（3）筛除风机不在满发状态运行的数据点

（4）筛除实际功率在理论功率1.2倍以上的数据点

最终得到的数据以csv格式输入。

2.模型建立：我们建立了五种在风功率预测表现良好的神经网络模型，分别是：RNN 循环神经网络，MLP 人工神经网络，LSTM 长短期记忆递归神经网络，

GRU 门控循环单元，CNN 卷积神经网络。并使用python建立对应的模型。

3.误差分析：我们引入神经网络模型的四种度量指标：

MAE（平均绝对误差），RMSE（均方根误差），MAPE（平均绝对百分比误差）和R2（决定系数）衡量模型精确度。

1. 预测功能：我们采用滑动窗口的时间序列预测方法实现对风功率的短时预测。
2. 界面开发：我们使用Qt Designer作为开发工具设计软件，要求能实现全部预期功能，同时具有良好的人机交互性。

# 4 功能与算法设计

## 4.1 功能设计

##### **4.1.1 数据读取与筛除**

数据是程序的基础。我们支持“普通运行数据”、“风机额定数据”、“直接训练数据”三种数据的导入。数据存储格式为csv，该格式优点为读写速度快、识别性强。

筛除是程序的保障。我们提供常规筛除、聚类筛除两种筛除方法。常规筛除速度快，用于筛除明显误差数据项；聚类筛除耗时较长，用于模型精细化处理。

##### **4.1.2 模型训练与误差分析**

模型训练是程序的核心。我们提供RNN、MLP、LSTM、GRU、CNN五种模型供用户选择，针对不同情况选择不同模型有利于精度的提高。

误差分析是程序的检验。机器学习的意义在于一层层网络推进，一步步迈向更高精度。误差分析给用户直观得到训练过程的状态变化，用于分析机器学习参数的选取区间。

##### **4.1.3 模型预测**

模型预测是程序的意义。由于风电功率预测与时间高度相关，我们选择“滑动窗口算法”，这要求我们预测内容与训练内容在时间上连续。因此我们直接在训练结果的后面直接绘制预测曲线，保持时间完整性。

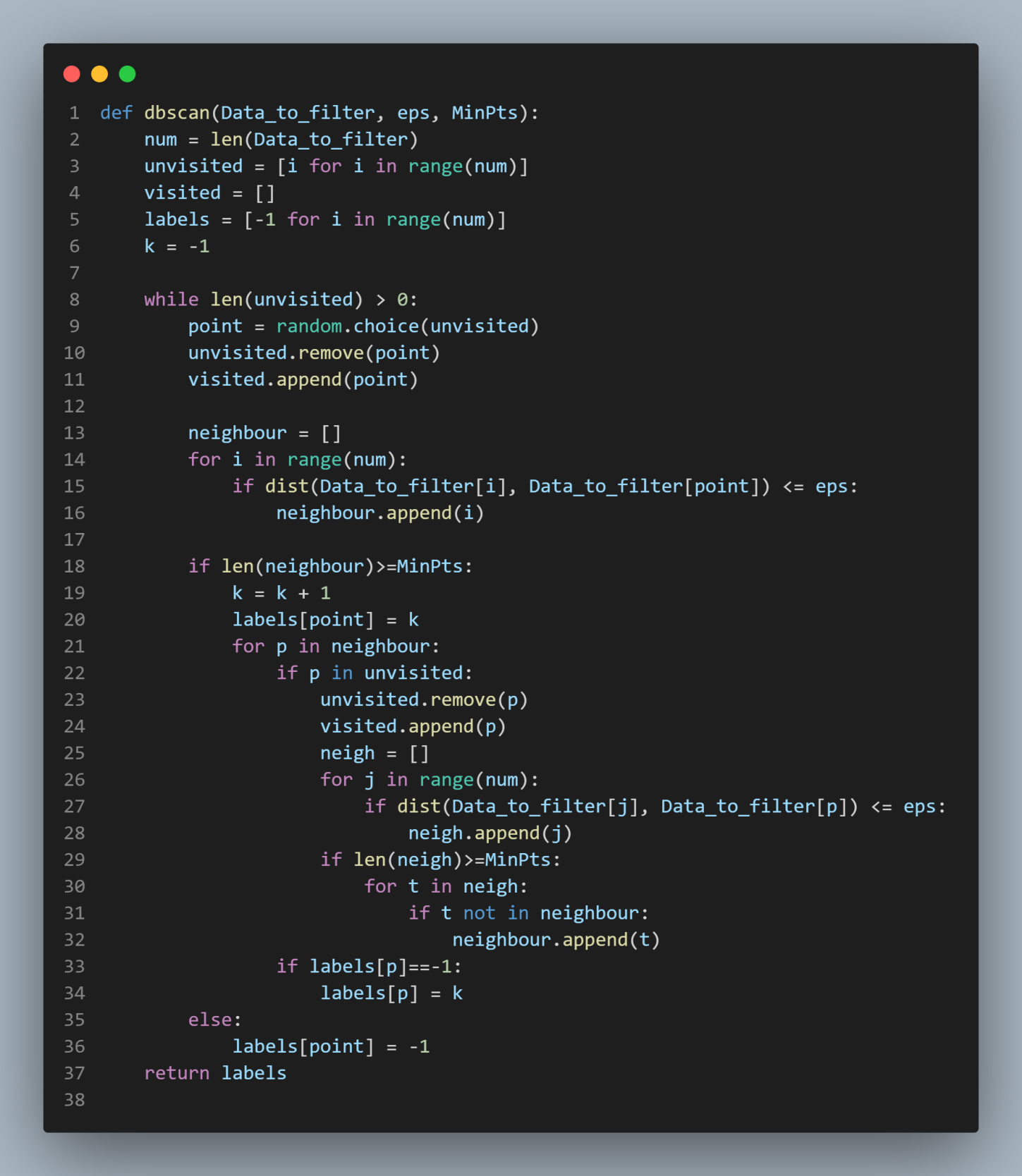
##### **4.1.4 高度自由化的参数设定**

## 4.2 核心算法

##### **4.2.1 聚类筛除算法**

由于风机正常运行的功率相对接近，我们采用基于密度的聚类算法dbscan，所得有效数据仅为一类，其余均为噪声点，能高效筛除掉无效数据。

由于scikit-learn库中提供的dbscan算法较为落后，其距离矩阵的占用内存极大且有效数据量小，实际运行中经常因内存不足而报错。我们手动计算其距离，并仅保留有效数据，以点为单位存储相对关系，大大缩小了内存占用。visit与unvisit保证了单点进搜索一次，大大缩减了聚类时间。



##### **4.2.2 滑动窗口算法**

由于风机运行功率变化与时间高度相关，我们采用了滑动窗口算法，将一段时间的数据打包作为单点的训练数据，如此获得的训练模型对时间高度敏感，能有效预测接下来一段时间的运行数据。滑动窗口算法是风电功率预测这样的时间连续性模型不二之选。



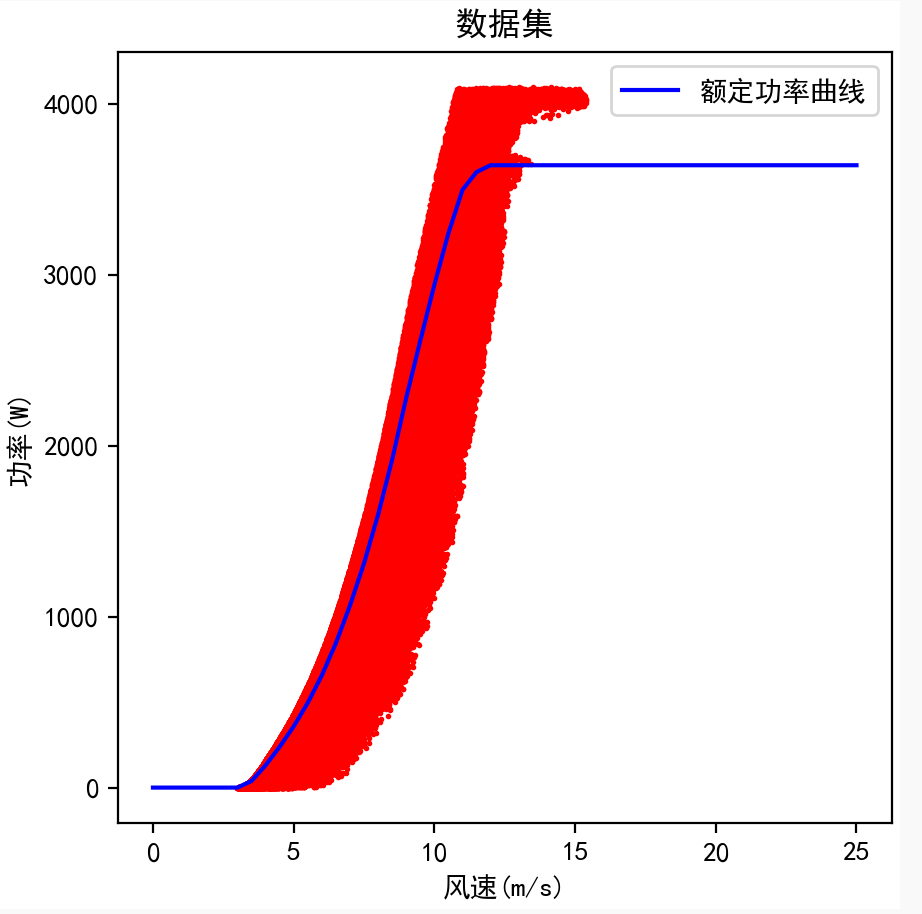
##### **4.2.3 机器学习模型建立**

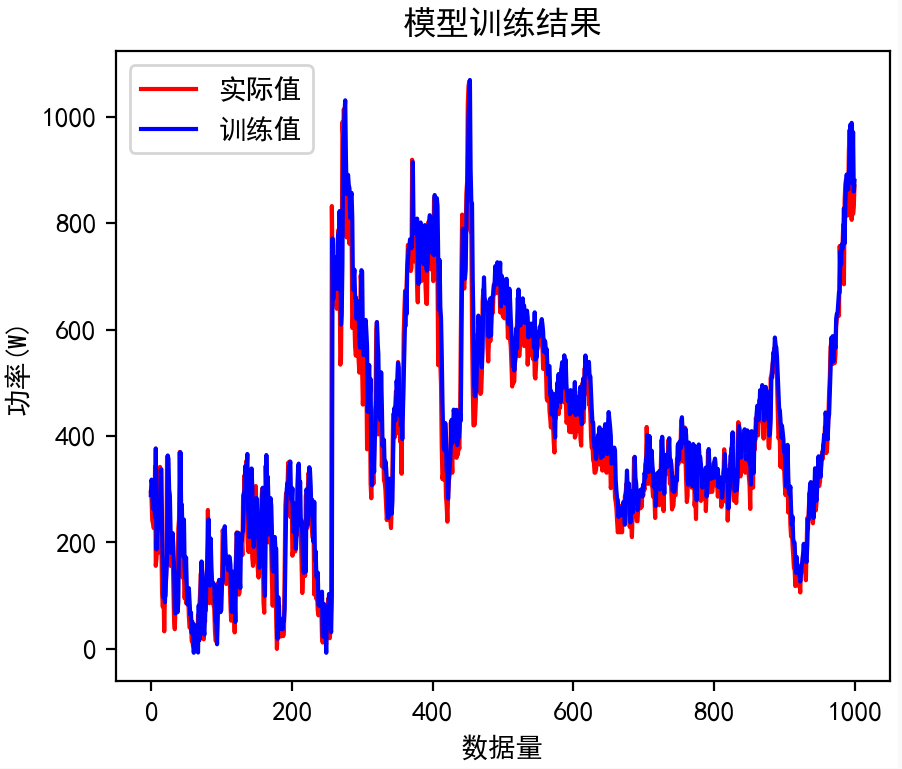
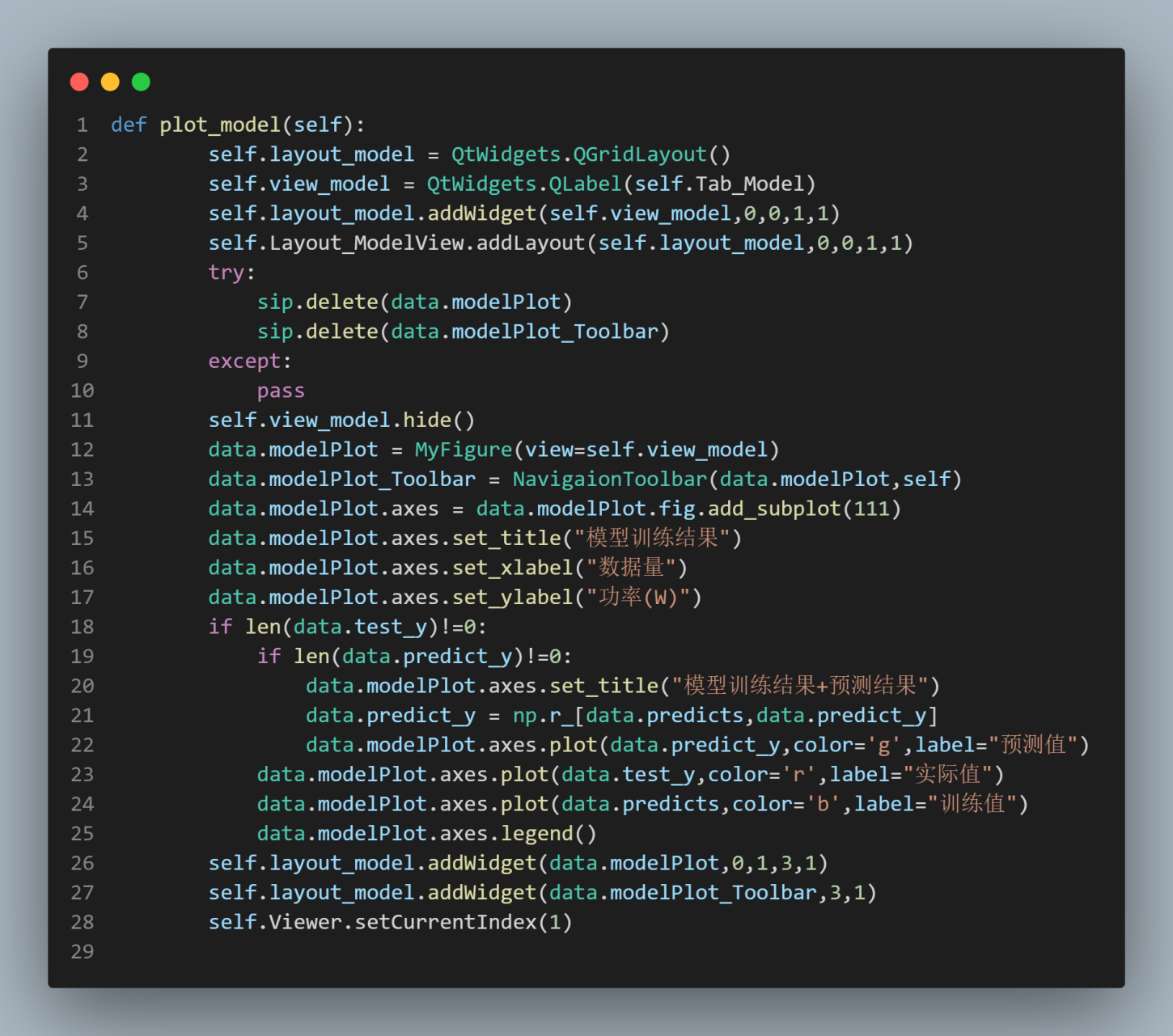
机器学习的关键在于模型的建立，我们采用了五种常见的机器学习模型——循环神经网络RNN、多层感知机MLP、长短期记忆神经网络LSTM、门控循环单元网络GRU、卷积神经网络CNN，给予用户多元化选择。

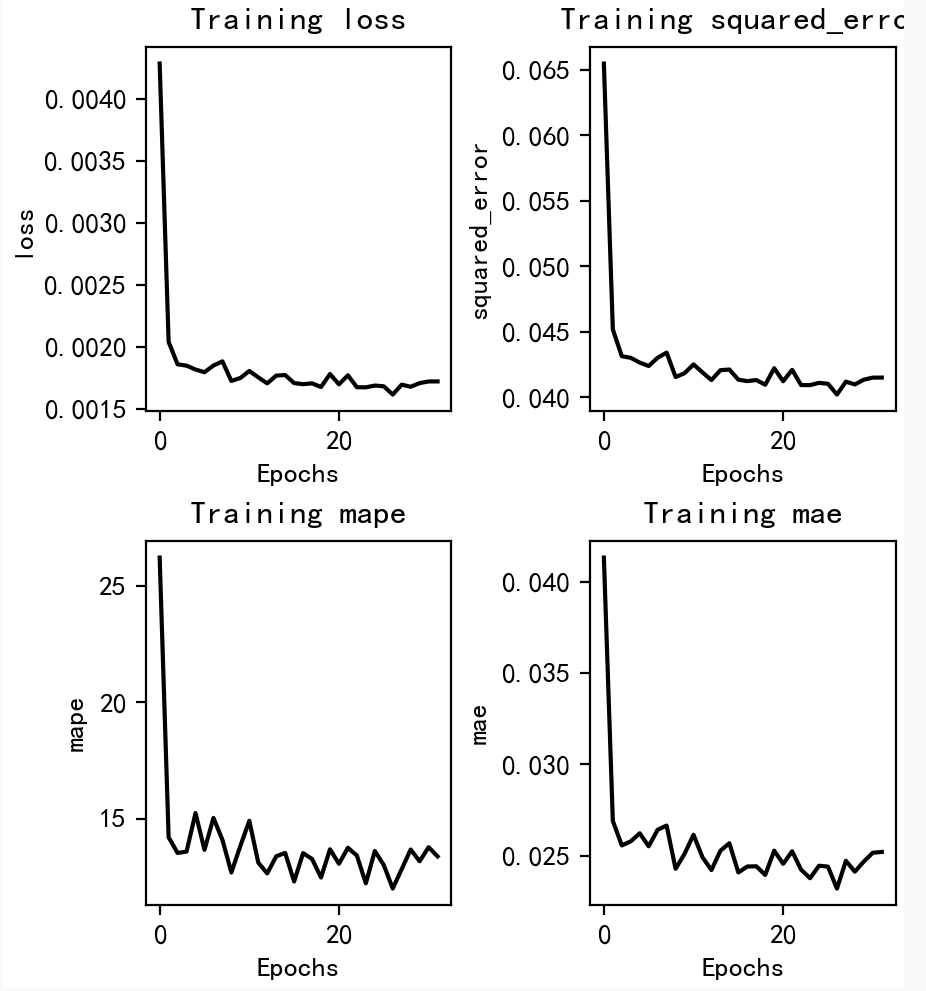


##### **4.2.4 全流程可视化支撑**

一款合格的软件应当具备全流程即时性可视化的特性。为此，我们设置“数据集”、“模型训练”、“误差分析”三个大板块各一张曲线图/散点图，覆盖风电功率预测全流程。读取数据、筛选数据，其数据的范围将在“数据集”中清晰展示。训练的模型与实际值的比较、未来变化曲线的预测在“模型训练”中可以直观看到。模型误差的评估、各误差指标随训练次数不断增加的变化情况也可以在“误差分析”中即时获取。







# 5 实施与运行

## 5.1 实施过程

## 5.2 运行效果评价

# 6 总结与反思

## 6.1 系统方案技术经济性评估

## 6.2 项目研究主要结论

## 6.3 系统适用场景分析与未来改进方向

## 6.4 课程学习过程反思