老师好，今天我们组结题答辩的项目是风电功率可视化软件开发。

我的答辩将从一下几个部分进行，分别是：选题意义，设计方案，技术路线，人员分工以及工程训练营总结

风电作为新能源发电的一种占比较高，且具有随机性波动性的特点，故实现对风电功率的精确预测可以有效解决风电波动性带来的运营和稳定性难题，协助电力系统实现智能调度，保障电网平稳运行，对于减少温室气体排放、推动能源可持续发展具有积极影响。

在项目开始之前，我们明确了整个开发过程的设计方案。在多技术比较中，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。

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。

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。

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，并通过常规和聚类的方法对数据进行筛除：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。

4、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。

5、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。

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列表保证了单点进搜索一次，大大缩减了聚类时间。

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。

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。

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。

我们的对模型的效果进行了评价，以测试数据为例，我们从104万个数据中选取了100000个数据进行模型训练，其中80000个为训练集，20000个为测试集。

随着训练次数的增加，我们可以看到误差指标整体呈下降趋势，整体拟合效果较好，基本达到了预测需求。

但是我们也看到，本次模型训练时间较长，为两小时8分钟，我们需要进一步优化算法，以提高模型训练效率。

曲嘉骏负责数据处理和搭建深度学习模型和部分的UI设计

李俊昊负责软件整体框架与UI交互界面设计，同时利用编程的优势在开发的各个部分都做出了贡献。

董启翰负责模型测试与预测精度提升

王乐天负责收集数据与制作答辩内容

在小组中，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。

经过全面系统评估，得出项目在技术与经济方面可行性与可优化的内容：

目标设定：准确预测未来短期内的风电出力

技术评估：数据质量高，模型精度种类多，精度较好，可视化界面设计优良

经济评估：项目开发周期为两周，软件功能多，前景好，也可进行进一步优化。

6.2 项目研究主要结论

经过技术经济性的可行性分析，得出以下结论：

1. 通过多种且合理的筛除方法能得到高质量数据，有利于提升模型精度

2、不同模型在处理不同数据集时表现不同，通过调整参数能得到更加有效的模型

3、良好的UI界面可以直观反映模型性能，提高工作效率

在软件工程训练营的学习过程中，我们获得了丰富的经验和知识，不仅在技术方面取得了进步，还在团队协作、问题解决和持续学习等方面有了深刻的认识。

6.3.2 未来改进方向：

1、风电场是风机的集合，空间尺度的扩大对风电功率预测产生误差，受制于地形等因素的影响较大，对风机集群功率的预测有必要进行。

2、训练层数需要自己设置，设置模型训练层数的深度学习，使得训练层数最优化。

3、获取更多历史数据，长期预测风电功率。使用其他方法，进行时间序列的单点预测。

4、集合神经网络模型四种度量指标，赋予权值，具体直观地将误差系数呈现在交互界面，供用户选择最适合的模型。