大家好，我今天要介绍的是《**社交媒体 多模态数据 的 统计挖掘 与 抑郁倾向 预测研究** 》

# 目录

四个点

# 研究背景

抑郁症影响着全球3.8亿人，但传统的诊断方法准确性较低，依赖人工评估，存在滞后性和主观性。同时，社交媒体用户每天产生大量的行为数据，包括文本内容、发帖时间、表情使用模式和社交互动信息，这些数据不仅记录了社交轨迹，也可能隐藏着情绪变化的信号。然而，以往的研究多采用单一模态进行分析，难以全面刻画抑郁倾向。

# 研究目的

因此，我们希望构建多模态融合模型，整合不同数据维度，提高预测准确性，并确保模型在心理学应用上的可解释性。

# 数据来源

基于这一研究目标，我们收集了15-17年国内外社交媒体平台的公开数据，其中包含用户的发帖文本、时间戳、表情符号和社交互动，并根据专业评分标准进行标注。这样的数据标注方式不仅提高了数据的医学可靠性，还结合了用户长期的社交行为特征，使模型能学习更具代表性的抑郁倾向模式。

# 数据预处理

为了保证数据质量，我们采用了**哈希去重**避免内容重复，文本中的敏感信息使用**正则表达式**匹配删除，同时对缺失数据进行填充，其中数值数据采用**多重插补法**，文本数据则通过Google api转换保持语义一致，并确保模型能够处理不同语言环境下的表情符号与情绪关键词。

# 多模态数据构成

这是我们的数据结构展示

# 文本模态处理

社交媒体文本表达往往带有情感色彩，抑郁用户的语言模式更倾向于负面情绪，因此，我们通过情感分析模型计算文本的情绪极性，发现抑郁用户的情感得分显著低于非抑郁组。同时，我们采用**融合模型**，确保模型既能捕捉复杂语义，又能通过词频信息提高可解释性。

# 时间模态处理

昼夜节律紊乱是抑郁的重要表征之一，研究发现抑郁用户在23:00-06:00的发帖频率显著高于正常群体，同时计算**夜间发帖占比**和**相邻发帖时间差**，进一步验证了**夜间活跃度**与**抑郁倾向**的正相关性。

# 表情模态处理

社交媒体中的表情符号能够反映用户的情绪状态，研究采用Emoji词典对表情符号进行分类，并计算**消极表情占比**，发现**抑郁用户**更倾向于使用负面表情，如“大哭”和“叹气”，**表情模态**的加入让模型能从**非文本化**情绪表达中提取抑郁倾向的信号。

# 社交数值模态处理

抑郁通常伴随**社交退缩**，因此研究分析了用户的社交互动数据，对**关注者数**、**转发数**和**点赞数**、**发帖数**进行**Z标准化**，并计算**关注-粉丝比**与**社交活跃度**，发现抑郁用户的**社交网络萎缩**是抑郁倾向的**重要预测因子**。

# 多模型融合模型结构设计

为了让不同模态的特征有效协同作用，研究采用**多头注意力机制**，通过**特征动态加权**优化**文本**、**时间**、**表情**和**社交数据**的融合，使模型不仅能学习**各模态特征**，还能**自动调整权重**，提高预测效果。

# 模型检验

通过评估模型，可以看出该模型拥有高准确率，远超单模态检查。此外，通过**损失函数的下降趋势**，可以看出模型在学习过程中有效捕捉了数据模式。

# 模型特征

研究发现抑郁用户在社交媒体上的行为表现有明显特征，**夜间发帖**增加了抑郁风险，**负面表情**与**负面语言**高度相关，**社交退缩**特征明显，这些多模态特征组合在一起加强了模型的预测能力。

# 研究总结

本研究突破了单模态分析的局限性，构建了基于文本、时间、表情和社交行为的多模态数据融合模型，实验表明该方法有效提高了**抑郁倾向识别**的准确性，同时分析验证了模型的心理学合理性，为**心理健康监测**提供了新的技术路径。

# 应用建议

社交媒体平台可以整合该预测模型，实现实时监测，在用户出现高风险行为时推送心理咨询，同时**优化推荐算法**，减少负面内容曝光，提高积极**信息推送**，在数据隐私保护的前提下，与医疗机构合作，为高风险用户提供心理健康咨询通道，推动心理学与计算机科学的深度融合。

# 结尾

最后，感谢各位评委老师！我的演讲到此结束。