

# 2026年o奖写作美赛指南

[一、O 奖核心逻辑与评审标准](#)

[二、可复用核心表格](#)

[三、全流程打卡表（还需修改，尤其是分工）](#)

[四、写作模板（逐句套用）](#)

[五、避坑指南（O 奖常见扣分点 + 案例避坑示例）](#)

[六、建模流程模板（步骤固定）](#)

[七、行文结构共性（写前必看始终遵循，不可直接gpt）](#)

[八、建模逻辑共性与其他注意（高分关键）](#)

目标：o奖

美赛：论文>代码

论文重中之重：摘要，论文图（用gemini3pro（nanobanana首选），echart（其次），origin复现优秀的图），对结果的讨论（雅思小作文），行文逻辑（这些是o）

参赛者：肖惠威，刘雨暄，孙浚涵

**论文评审标准：**

是否给出了令人满意的赛题解读，对赛题中的模糊概念是否给出了必要澄清；是否明确列出了建模需要的**前提条件及假设**，对其合理性是否给出了满意的解释或论证；是否根据对赛题的分析给出了建模的**动机或论证了建模的合理性**；是否设计出了**有效的解题模型**；是否对模型给出了**稳定性测试**；是否讨论了模型的**优缺点**，并给出了**清晰的结论**；是否给出了符合要求的**摘要**。

## ▼ 一、O 奖核心逻辑与评审标准

### 1. 核心逻辑（所有 O 奖论文通用）

- 破题精准：拆解赛题为 3-5 个可建模子任务，澄清模糊概念（如奥运论文定义“进步国家”为“奖牌数波动超 2 枚”，网球论文定义“动量”为“实时胜率 - 赛前胜率”）。
- 模型适配：拒绝盲目堆砌，优先“经典模型改良”（如 GSRF = 网格搜索 + 随机森林）或“多模型融合”（如 ARIMA-LSTM），每个模型必说明适配理由（如 Wordle 论文用 LSTM 适配时间序列预测，网球论文用纳什均衡适配战术策略分析）。
- 细节落地：数据预处理、参数优化、模型验证均有具体操作，结论需“数据 + 公式 + 图表”三重支撑（如奥运论文用 Spearman 相关系数 0.874 证明教练影响，网球论文用 K-S 检验  $p < 0.05$  验证动量显著性）。
- 逻辑闭环：任务间衔接自然（如奥运论文 Task1 的 GSRF 预测结果直接作为 Task3 “项目 - 奖牌关系”的输入），模型有评估、有稳健性测试、有创新点验证。

### 2. 美赛评审核心标准（附案例达标示例）

评审维度	核心要求	达标标准	O 奖案例示例
问题解读	明确任务边界，澄清模糊点	输出《赛题拆解表》，无歧义表述	网球论文将“动量”量化为“实时胜率与赛前胜率差值”，排除发球优势干扰
模型质量	适配任务、参数优化、稳健性强	核心指标达标（回归 $R^2 \geq 0.7$ , 分类 $AUC \geq 0.8$ ），灵敏度波动 $\leq 1.5\%$	奥运论文 GSRF 模型 $R^2 = 0.784$ ，灵敏度分析波动仅 0.058%；Wordle 论文 Stacking 模型 $R^2 = 0.8377$
数据支撑	预处理规范，特征工程合理	异常值 / 缺失值处理有依据，特征筛选有方法（RFE / 因子分析）	网球论文用 Bartlett 球形度检验 ( $p < 0.05$ ) 验证因子分析合理性；Wordle 论文用 RobustScaler 处理异常值
创新价值	模型改良或新指标，可验证	创新点有定量指标支撑（如 $R^2$ 提升 0.29）	奥运论文提出“伟大教练模型”，罗马尼亚体操得分从 3 提升至 36.65；网球论文提出 RTMQ 实时动量量化模型
写作规范	逻辑清晰，格式统一	摘要浓缩核心，图表有标题 + 图例 + 编号，公式编号连续	所有 O 奖论文摘要均 200-300 词，含“问题→方法→关键结果→创新点”，图表编号连续且与正文对应

## ▼ 二、可复用核心表格

表 1：模型参数与适配性对照表（不用参考这个，具体更全的在腾讯文档知识库里）

任务类型	核心模型	关键参数 (O 奖案例示例)	适配性理由	优化方法	对应赛题案例	负责人	完成情况
预测类 (奖牌 / 玩家数)	GSRF	学习率 = 0.2、 max_depth=5、 n_estimators=300	网格搜索优化超参数, 抗过拟合	10 折交叉验证确定最优参数	奥运奖牌预测	建模组	□
分类类 (获奖国家 / 单词难度)	逻辑回归	正则化强度 $\lambda=0.1$ 、迭代次数 = 1000	适配二分类问题, 计算高效	梯度下降最小化交叉熵损失	奥运未获奖国家预测	建模组	□
策略类 (网球战术)	混合策略纳什均衡	服务器策略数 = 4、接收方策略数 = 3	覆盖攻防全场景, 贴合真实战术选择	K-means 聚类划分策略集群	网球动量分析	建模组	□
时间序列 (Wordle 玩家数)	ARIMA-LSTM	ARIMA ( $p=2, d=1, q=0$ )、 LSTM 隐藏层 = 20	线性 + 非线性结合, 捕捉趋势与波动	残差互补优化预测精度	Wordle 游戏建模	建模组	□

表 2：数据质量校验表（参考部分，具体更全的在腾讯文档知识库里）

数据环节	校验项	校验方法	合格标准	不合格处理方案	对应赛题案例	负责人	完成情况
数据清洗	异常值检测	箱线图 (IQR 法则)	异常值占比≤3%	替换为中位数 / 删除异常样本	Wordle 文论剔除非 5 字母词	数据组	□
缺失值处理	关键特征缺失率	统计各列缺失值占比	核心特征缺失率≤5%	均值填充 (数值型) / 众数填充 (分类型)	网球论文用平均球速填充缺失值	数据组	□
特征工程	特征相关性	皮尔逊 / 斯皮尔曼相关系数	冗余特征 ( $r \geq 0.8$ ) ≤2 个	剔除高相关特征	奥运论文剔除运动员数与项目数高相关特征	数据组	□
数据标准化	特征量纲一致性	查看均值 / 标准差	标准化后均值≈0、标准差≈1	重新执行 Min-Max/Z-score 标准化	Wordle 文论用 RobustScaler 处理词频数据	数据组	□

表 3：创新点验证表（先了解领域前沿成果和其他团队解法，结合已知模型解决问题。不要拿到题目就套用已有模型，比赛时应花半天时间查找和阅读相关文献最好有新的模型（但不是首要优先级），这里附案例填写）

创新类型	创新内容	验证指标	目标值 (O 奖水平)	验证方法	对应赛题案例	负责人
模型改良	GSRF (网格搜索 + 随机森林)	较基础模型 $R^2$ 提升幅度	≥0.2 (ARIMA $R^2=0.494 \rightarrow GSRF=0.784$ )	同一数据集对比基础模型指标	奥运奖牌预测	建模组
新指标创建	动态竞赛指数 (DPI)	与真实比赛趋势契合度	≥85% (贴合网球比分变化)	可视化对比指标与实际结果	网球动量分析	写作组
方法融合	ARIMA-LSTM 残差互补	预测区间覆盖率	95% 置信区间覆盖真实值 ≥90%	扰动输入变量观察预测稳定性	Wordle 玩家数预测	建模组
场景适配	多场地网球模型	跨场地预测精度下降幅度	≤15% (温网→法网精度下降≤10%)	不同场景数据集交叉验证	网球动量分析	建模组

表 4：成果可视化清单表（还需拓展主要用 nano-banana，这地方等可视化效果很多，附案例填写）

可视化类型	核心用途	推荐图表 (O 奖高频)	关键要素 (评审关注点)	完成标准	对应赛题案例	负责人	完成情况
特征重要性	凸显关键影响因素	雷达图 / 热力图	标注特征贡献率 (如运动员数占比 74.1%)	图表含标题 + 图例 + 数值标签	奥运论文用雷达图展示运动员数、项目数对奖牌的影响	写作组	□
预测结果对比	验证模型精度	误差棒图 / 真实值 vs 预测值折线图	标注 MAE/RMSE 等核心指标	预测值与真实值偏差≤10%	Wordle 文论对比 LSTM 与 ARIMA 预测结果	写作组	□
分类效果展示	体现分类准确性	混淆矩阵热力图 / ROC 曲线	标注 AUC/F1-score ( $\geq 0.8$ )	曲线光滑、阈值明确	奥运未获奖国家分类 (AUC=0.918)	写作组	□

可视化类型	核心用途	推荐图表（O 奖高频）	关键要素（评审关注点）	完成标准	对应赛题案例	负责人	完成状态
创新点对比	量化创新价值	柱状图（基础模型 vs 改进模型）	明确标注指标提升幅度（如 $R^2 + 0.29$ ）	对比维度统一、数据标签清晰	网球论文对比 RTMQ 模型与传统恒定胜率模型	写作组	<input type="checkbox"/>

### ▼ 三、全流程打卡表（还需修改，尤其是分工）

#### 第一阶段：前期准备（0.5天到1天）

序号	任务名称	必做事项	验收标准	对应 O 奖案例做法	负责人	完成状态
1	赛题拆解	1. 逐句精读赛题，明确核心任务（预测 / 分类 / 策略）； 2. 澄清模糊概念（如“动量”定义）； 3. 拆解为 3-5 个可建模子任务，明确任务间逻辑关系	输出《赛题拆解表》，含“任务描述 + 核心目标 + 输入输出数据”	奥运论文拆解为“奖牌预测、未获奖国家分类、项目 - 奖牌关系、教练影响、原创见解”5 个任务	全员	<input type="checkbox"/>
2	假设与符号定义	1. 列出 3-5 个合理假设，每个假设附数据 / 现实支撑； 2. 统一符号体系，用表格呈现（符号 + 定义 + 单位）	假设无歧义，符号前后一致，覆盖模型核心变量	Wordle 论文假设“用户报告分数真实”，理由是“Twitter 数据经官方筛选，异常值占比 < 3%”	建模组	<input type="checkbox"/> 1
3	数据预处理	1. 清洗异常值（剔除 / 替换）、缺失值（填充 / 剔除），说明处理理由； 2. 标准化数据（Min-Max/Z-score）； 3. 特征工程（筛选核心特征，剔除冗余）	输出《数据预处理报告》，含处理步骤 + 处理后数据样本	网球论文用 ISO 映射统一国家名称，剔除中立运动员数据；Wordle 论文剔除非 5 字母词和百分比求和≠100% 的数据	数据组	<input type="checkbox"/> 2
4	模型选型	1. 按子任务类型选适配模型（预测 → GSRF/LSTM；分类 → Logistic/SVM；策略 → 纳什均衡）； 2. 说明模型适配理由（如“LSTM 适配时间序列，抗长距离依赖”）	输出《模型选型表》，含“任务 + 模型 + 适配理由”	奥运论文 Task1 用 GSRF（抗过拟合），Task2 用逻辑回归（适配二分类），Task4 用 Lasso 回归（筛选关键特征）	建模组	<input type="checkbox"/>

#### 第二阶段：建模执行（第 3-6 天）

序号	任务名称	必做事项	验收标准	对应 O 奖案例做法	负责人	完成状态
1	模型构建	1. 推导核心公式，标注符号含义；2. 优化参数（网格搜索 / 交叉验证）；3. 代码实现，保留关键代码片段	公式推导逻辑清晰，参数有具体数值（如“GBDT 学习率 = 0.2”），代码可复现	网球论文用 10 折交叉验证确定 LSTM 隐藏层 = 20；Wordle 论文用网格搜索确定 XGBoost 最优参数	建模组	<input type="checkbox"/>
2	模型验证	1. 定量评估（回归用 $R^2$ /MAE/RMSE；分类用 AUC/F1/混淆矩阵）； 2. 稳健性测试（扰动关键特征，观察模型变化）	核心指标达标（如 $R^2 \geq 0.7$ , $AUC \geq 0.8$ ），波动范围 $\leq 1.5\%$ ，泛化测试精度无显著下降	奥运论文扰动运动员数，模型波动 0.058%；网球论文用温网 + 法网数据验证泛化能力	建模组	<input type="checkbox"/>

序号	任务名称	必做事项	验收标准	对应O奖案例做法	负责人	完成状态
		型波动)； 3. 泛化测试（不同数据集 / 场景验证）				
3	结果可视化	1. 用雷达图 / 折线图 / 热力图展示核心结果（如特征重要性、预测值 vs 真实值）； 2. 图表含“标题 + 图例 + 编号”，与正文对应	图表简洁直观，能直接支撑核心结论（如“雷达图显示运动员数是奖牌预测关键特征”）	奥运论文用误差棒图展示 2028 奖牌预测区间；网球论文用傅里叶拟合图展示动量趋势	写作组	<input type="checkbox"/>
4	创新点落地	1. 实现 1-2 个创新点（模型改良 / 新指标）； 2. 用定量指标验证创新价值（如“GSRF 比 ARIMA R <sup>2</sup> 提升 0.29”）	创新点有明确公式 / 代码支撑，验证结果可量化	网球论文提出 RTMQ 模型，通过 K-S 检验验证其有效性；Wordle 论文提出 MIMO XGBoost，预测精度提升 12%	建模组	<input type="checkbox"/>

### 第三阶段：写作优化（第 7-9 天）

序号	任务名称	必做事项	验收标准	对应O奖案例做法	负责人	完成状态
1	框架搭建	1. 按“摘要→引言→模型→结果→结论→附录”排版； 2. 每个章节用“问题→模型→结果→分析”逻辑链串联； 3. 插入核心流程图（研究路径 / 模型架构）	框架符合 O 奖标准，章节衔接自然，无孤立内容	所有 O 奖论文均含“Our Work”流程图，清晰展示“数据→模型→任务→结论”路径	写作组	<input type="checkbox"/>
2	正文写作	1. 按模板撰写关键段落（问题重述 / 模型介绍 / 结果分析）； 2. 每个结论配“数据 + 公式 + 图表”支撑； 3. 避免空泛描述，用定量指标说话（如“R <sup>2</sup> =0.784, MAE=0.914”）	正文逻辑闭环，无冗余文字，数据 / 公式 / 图表对应一致	奥运论文用公式 + 雷达图 + 相关系数证明教练对奖牌的影响；Wordle 论文用混淆矩阵 + ROC 曲线展示分类效果	写作组	<input type="checkbox"/>
3	原创见解与建议	1. 基于模型结果给出落地建议（如给 IOC / 教练 / 纽约时报的具体方案）； 2. 提出 1-2 个原创观点（如“东道主奖牌数提升 1.4-1.9 倍”）	建议可操作，原创观点有模型结果支撑，不脱离赛题	奥运论文建议 IOC 鼓励国家挖掘天才运动员；Wordle 论文建议新增难度分级机制；网球论文给教练提供战术调整方案	全员	<input type="checkbox"/>
4	参考文献与附录	1. 引用 6-10 篇相关文献（APA 格式）； 2. 附录含“核心代码 + 补充数据 + 公式推导细节”	参考文献格式统一，附录内容完整，可支撑正文结论	所有 O 奖论文附录含完整代码片段、补充数据表格、公式推导步骤，参考文献涵盖相关领域经典研究	写作组	<input type="checkbox"/>

### 第四阶段：终审检查（第 10 天）

序号	检查项目	必查内容	验收标准	对应O奖案例合格示例	检查人	合格状态
1	模型质量检查	1. 核心指标是否达标 ( $R^2 \geq 0.7$ )，	模型无逻辑漏洞，指标优于基	网球论文 SP 模型预测精度	建模组组长	<input type="checkbox"/>

序号	检查项目	必查内容	验收标准	对应O奖案例合 格示例	检查人	合格状态
		AUC≥0.8， MAE/RMSE 较 小)； 2. 稳健性 / 泛化测 试是否完成； 3. 创新点是否验证	基础模型，创新点 有量化支撑	90.44%，跨场地 泛化精度下降 ≤10%		
2	写作规范检查	1. 摘要 (200-300 词) 含 “问题→方法 →关键结果→创新 点”； 2. 图表有标题 + 图 例 + 编号，公式编 号连续； 3. 符号前后一致， 无语法错误	摘要浓缩核心， 格式规范，无细 节漏洞	奥运论文摘要直接 给出 GSRF 的 $R^2=0.784$ 、预测 区间，无冗余文字	写作组组长	<input type="checkbox"/>
3	避坑项检查	1. 无无依据假设； 2. 无模型堆砌 (每 个模型都有适配理 由)； 3. 无数据 / 公式 / 图 表矛盾；4. 创新点 有验证	避开 O 奖常见扣 分点，符合 “精准 破题 - 模型适配 - 细节落地” 逻辑	网球论文每个模型 都说明适配性 (如 傅里叶拟合适配动 量趋势提取)，无 冗余模型	全员	<input type="checkbox"/>
4	格式合规检查	1. 总页数≤25 页 (含附录)；2. 字体 / 行距统一，页码连 续； 3. 参考文献格式正 确	符合美赛格式要 求，无格式混乱 问题	所有 O 奖论文均 控制在 25 页内， 图表编号连续，参 考文献 APA 格式 统一	写作组	<input type="checkbox"/>

## ▼ 四、写作模板（逐句套用）

### 1. 问题重述段

“针对赛题 XX 任务，本文首先澄清 XX 模糊点（定义 XX 概念），随后拆解为 X 个子任务，分别采用 XX 模型（改良自 XX）解决，最终通过 XX 指标验证模型有效性，给出 XX 落地建议。”

- 案例套用：“针对赛题网球动量分析任务，本文首先澄清‘动量’模糊点（定义为‘球员实时胜率与赛前胜率的差值，排除发球优势干扰’），随后拆解为‘动量量化、动量显著性验证、动量转向预测、战术策略优化’4 个子任务，分别采用 RTMQ 模型（改良自 Elo 评分系统）、K-S 检验、多元线性回归、混合策略纳什均衡模型解决，最终通过  $AUC=0.918$ 、预测精度 90.44% 验证模型有效性，给出给教练的战术调整建议。”

### 2. 模型介绍段

“为解决 XX 问题，选择 XX 模型，其核心优势是 XX（如适配非线性数据 / 抗过拟合）。模型输入为 XX 特征，通过 XX 方法（网格搜索 / 交叉验证）优化参数，核心公式为 XX，最终实现 XX 效果。”

- 案例套用：“为解决 2028 奥运奖牌预测问题，选择 GSRF 模型，其核心优势是通过网格搜索优化超参数，有效缓解过拟合。模型输入为‘运动员数、历史奖牌数、赛事总数、东道主身份’4 个特征，通过 10 折交叉验证确定最优参数（学习率 = 0.2、 $\max\_depth=5$ 、 $n\_estimators=300$ ），核心公式为  $Yg=GSRF(Aij,g,Gij,Ej,g,Hij,g)$ ，最终实现预测  $R^2=0.784$  的效果。”

### 3. 结果分析段

“模型验证显示（表 X / 图 X），XX 指标达 XX（如  $R^2=0.784$ 、 $MAE=0.914$ 、 $AUC=0.918$ ），说明 XX 结论；进一步分析表明，XX 因素对结果影响最大（如运动员数贡献 74.1%），与 XX 现实逻辑一致。”

- 案例套用：“模型验证显示（表 3 / 图 6），GSRF 模型总奖牌预测  $R^2=0.784$ 、 $MAE=2.430$ ，说明模型预测精度较高；进一步分析表明，运动员数对奖牌预测贡献 74.1%，与‘更多运动员意味着更广选拔范围’的现实逻辑一致。”

### 4. 模型优劣段

“本模型的优势在于：1) XX（如 GSRF 用超参数优化提升精度）；2) XX（如灵敏度分析验证稳健性）；3) XX（如结果可视化清晰直观）。局限性包括：1) XX（如假设变量独立，未考虑交互效应）；2) XX（如未涵盖极端场景影响）。”

- 案例套用：“本模型的优势在于：1) GSRF 模型通过网格搜索优化超参数，较传统 ARIMA 模型 R<sup>2</sup> 提升 0.29；2) 灵敏度分析显示，扰动运动员数后模型波动仅 0.058%，验证稳健性；3) 用雷达图、误差棒图直观展示核心结果，便于评审理解。局限性包括：1) 假设 2028 年参赛数与 2024 一致，未考虑实际变化；2) 未考虑运动员数与赛事数的正相关交互效应，可能导致预测偏差。”

## 5. 摘要模板（200-300 词）

“针对 XX 赛题要求，本文围绕 XX 核心目标，拆解为 X 个子任务。首先通过 XX 方法预处理数据，随后构建 XX 模型（改良自 XX）解决 XX 问题，通过 XX 模型解决 XX 问题。模型验证显示，XX 指标达 XX（如 R<sup>2</sup>=0.784、AUC=0.918），稳健性测试表明模型波动≤1.5%。创新点在于 XX（模型改良 / 新指标），其价值通过 XX 定量验证（如 R<sup>2</sup> 提升 0.29）。基于模型结果，提出 XX 落地建议（如给 XX 的具体方案），为 XX 提供参考。”

- 案例套用：“针对奥运奖牌预测赛题要求，本文围绕‘2028 奖牌榜预测、未获奖国家分类、项目 - 奖牌关系分析’核心目标，拆解为 5 个子任务。首先通过 ISO 映射统一国家名称、剔除异常值预处理数据，随后构建 GSRF 模型（改良自随机森林）解决奖牌预测问题，通过逻辑回归模型解决未获奖国家分类问题。模型验证显示，GSRF 模型 R<sup>2</sup>=0.784、逻辑回归 AUC=0.918，稳健性测试表明模型波动≤1.5%。创新点在于提出‘伟大教练模型’，其价值通过 Spearman 相关系数 0.874 定量验证。基于模型结果，提出 IOC 鼓励国家主办奥运、挖掘天才运动员的落地建议，为奥运赛事发展提供参考。”

## ▼ 五、避坑指南（O 奖常见扣分点 + 案例避坑示例）

1. 假设无依据：O 奖论文的假设均有数据 / 现实支撑，避免“假设变量无交互”等无理由假设。
  - 避坑示例：奥运论文假设“2028 奥运参赛数与 2024 一致”，理由是“近年参赛国数波动 < 5%”，而非单纯假设。
2. 模型堆砌复杂：拒绝为复杂而复杂，如短期预测用灰度模型、长期用 GRU，而非全程用深度学习，每个模型必说明“适配性”。
  - 避坑示例：Wordle 论文短期预测用灰度模型（适配小数据），长期用 LSTM（适配时间序列），而非全程用 Transformer。
3. 摘要冗余：必须浓缩“问题→方法→关键结果→创新点”，无废话（如直接给出核心指标和结论）。
  - 避坑示例：网球论文摘要直接给出 RTMQ 模型 K-S 检验 p<0.05、SP 模型预测精度 90.44%，无冗余背景描述。
4. 细节马虎：符号定义前后一致（如 O 奖论文中 A<sub>ij</sub>始终表示“i 国 j 届奥运运动员数”）、公式推导无漏洞、表格数据无矛盾。
  - 避坑示例：所有 O 奖论文均用表格统一符号定义，公式编号连续，预测百分比总和接近 100%。
5. 创新无验证：所有创新点（改良模型、新指标）均需用定量指标验证。
  - 避坑示例：网球论文提出 RTMQ 模型后，用 K-S 检验 p<0.05 证明其有效性，而非单纯描述模型结构。

## ▼ 六、建模流程模板（步骤固定）

### 1. 数据处理

1. 清洗：处理异常值（剔除 / 替换）、缺失值（填充 / 剔除），说明处理理由（如 Wordle 论文剔除非 5 字母词，理由是“不符合游戏规则”）；
2. 标准化：根据数据特征选择 Min-Max/Z-score/RobustScaler，避免特征权重失衡（如奥运论文用 Z-score 标准化运动员数、赛事数）；
3. 特征工程：通过 RFE / 因子分析 / 相关性分析筛选核心特征，剔除冗余变量（如网球论文用因子分析从 25 个指标中提取 5 个核心因子）。

### 2. 模型构建

1. 任务拆解：将复杂赛题拆分为具体可建模子任务（如奥运论文拆为 5 个任务，Wordle 论文拆为 4 个任务）；
2. 模型选择：根据任务类型选适配模型（预测→GSRF/LSTM/ARIMA-LSTM；分类→Logistic/SVM/GMM；策略→纳什均衡）；
3. 参数优化：用网格搜索 / 交叉验证确定最优参数（如 GBDT 最优参数：学习率 0.2、max depth=5、n\_estimators=300）；
4. 公式推导：明确模型核心公式，标注符号含义，推导逻辑清晰（如 LSTM 的遗忘门、输入门公式，纳什均衡的期望收益公式）。

### 3. 验证环节

- 定量评估：用  $R^2$ 、MAE、RMSE（回归），AUC、混淆矩阵、F1-score（分类）验证模型精度（如奥运论文 GSRF 模型  $R^2=0.784$ , Wordle 论文 SVM 分类  $F1=0.6634$ ）；
- 稳健性测试：扰动关键特征（如运动员数、词频），观察模型输出波动，确保波动范围小（如平均灵敏度 < 1%）；
- 泛化测试：用不同数据集 / 场景验证（如网球论文用温网 + 法网数据，Wordle 论文用不同难度单词验证）。

#### 4. 结论输出

- 核心发现：提炼每个任务的关键结论（如“中美仍将领跑 2028 奥运奖牌榜”“ERIE 是 Wordle 难词”）；
- 落地建议：结合结论给出可操作建议（如“IOC 可鼓励国家申请主办奥运”“Wordle 可新增难度分级机制”）；
- 局限与改进：客观指出模型不足（如“未考虑变量交互”“未涵盖极端天气影响”），提出改进方向。

### ▼ 七、行文结构共性（写前必看始终遵循，不可直接gpt）

#### 1. 开篇三段式（定调 + 破题 + 路径）

- 背景引入：1-2 段结合现实场景（如奥运奖牌竞争、网球赛事翻盘、Wordle 游戏热度），引出问题价值，不冗余。
- 问题重述 + 澄清：逐任务拆解赛题，明确模糊概念（如“动量”定义为“实时胜率 - 赛前胜率”、“进步国家”定义为“奖牌数波动超 2 枚”），避免歧义。
- 核心路径图：用可视化流程图展示“数据→模型→任务→结论”全流程（如“Our Work”图、“Model Overview”图），让评审一眼看懂研究逻辑。

#### 2. 中间主体（宏观数据 - 模型 - 验证，层层递进）

- 前置准备：先列“假设 + 符号定义”，假设必附合理性说明（如“2028 奥运参赛数与 2024 一致”，理由是“近年数据波动 < 5%”），符号用表格清晰呈现，前后一致。
- 数据预处理：明确“异常值处理 + 标准化 + 特征工程”步骤（如 Wordle 论文剔除非 5 字母词、奥运论文用 ISO 映射统一国家名称），每个操作说明“为何做”（如标准化避免特征权重失衡）。
- 分任务建模：每个任务遵循“模型选择理由→公式推导→参数优化→结果可视化”，模型选择不盲目（预测用 GSRF/LSTM、分类用 Logistic/SVM、策略用纳什均衡），必说明“适配性”。
- 验证环节：必做“模型评估 + 灵敏度分析”，评估用定量指标（ $R^2$ 、MAE、AUC、混淆矩阵），灵敏度分析通过“扰动关键指标”验证稳健性（如奥运论文扰动运动员数，模型波动 < 1.5%）。

#### 3. 中间主体（微观）

摘要原文

Introduction原文

Assumption&Notation原文

Data Collection and Preprocessing原文

第四部分模型，也就是文章主体部分要参考论文原文

Sensitivity and Robustness Analysis原文

模型优势与不足

目录章节	共性行文思路（固定逻辑）	O奖案例示例（以Wordle游戏建模赛题为例）
<b>Abstract</b>	<ol style="list-style-type: none"> <li>问题：赛题核心需求（以...为目标，本文旨在，为了）+做了啥（本文构建了，验证等）；</li> <li>方法：2-3个核心模型+基于什么背景和约束（参考，基于，结合）或+使用目的（为预测大幅动量波动，我们引入<b>波动预测（SP）模型</b>）；</li> <li>结果与结论：关键量化指标，对题目的回应；</li> <li>创新：模型/指标创新；（可选，属于后</li> </ol>	<ol style="list-style-type: none"> <li>问题：需预测Wordle玩家得分分布、分类单词难度；</li> <li>方法：ARIMA-LSTM（适配时间序列）、GMM（适配难度聚类）；</li> <li>结果与结论：ARIMA-LSTM预测 <math>R^2=0.8377</math>, GMM分类准确率67%；</li> <li>创新：用Stacking融合线性与树模型；</li> <li>价值：给纽约时报提供单词难度分级建议。</li> </ol>

目录章节	共性行文思路（固定逻辑）	O奖案例示例（以Wordle游戏建模赛题为例）
	面的迭代了）； 5. 价值：落地建议；	
1. Introduction	1. 背景：现实场景（Wordle热度）+问题价值（网球动量那个写得不错）； 2. 赛题拆解（问题澄清或重述）：澄清“单词难度”定义+拆分为3子任务，这个和题目里的给出的问题区别（在intro原文界面里）；也就是建模性的任务+具体的实现路径；是“建模”思维的体现； 3. 文献综述（可？）（Literature Review）：简述领域已有研究+本文创新点。 4. 我们的工作（一定需要流程图，nano banana）分框多少个task多少个框	1. 背景：Wordle日活超200万，但单词难度无统一标准，影响用户体验； 2. 赛题拆解：定义“难度”为“平均尝试次数”，拆分为“得分预测、难度分类、优化建议”； 3. 文献综述：已有研究用词频预测难度，本文加入“字母位置特征”，提升精度12%。
2. Assumptions and Notations	1. 假设：3-5个合理假设+数据支撑； 2. 符号：表格（符号+定义+单位）+前后一致。	1. 假设：“用户报告得分真实”（支撑：Twitter数据异常值占比<3%）； 2. 符号： $\$ATN_{-i\$}$ ：第 <i>i</i> 个单词平均尝试次数（次）； $\$F_{-ij\$}$ ：第 <i>i</i> 个单词第 <i>j</i> 个字母的词频（%）。
3. Data Collection and Preprocessing	1. 数据来源：说明渠道（如GitHub/API）； 2. 清洗：异常值/缺失值处理+理由； 3. 标准化：选择方法（如RobustScaler）+理由；	1. 数据来源：GitHub公开的2022-2023年Wordle得分数据； 2. 清洗：剔除“尝试次数>6”的数据（理由：游戏规则上限为6次）； 3. 标准化：用RobustScaler（理由：抗词频异常值）； 4. 特征工程：用RFE筛选出“词频、字母位置、重复字母数”3核心特征。
4. Model Development, result and Discussion (这里是论文的主体，建议多参考原文)	1 定义问题 2 特征工程：筛选核心特征+方法（如RFE选了哪些特征，原因，合理性）； 3 模型构建和训练的流程+选择理由+公式+参数优化； 4 模型结果和讨论：子任务结果：量化数据+图表（小表大图）；结果解读：每图/表配3行分析（数据问题+现实关联，这里就是雅思小作文了，很重要）； 5 对比分析（如有）：与基础模型对比+优势	1 先定义“用报告结果的加权平均来衡量单词难度”，再基于此定义制定分类标准，然后构建SVM模型，最后评估模型并预测“EERIE”。  2 特征工程：用RFE筛选出“词频、字母位置、重复字母数”3核心特征。  3. 见具体论文  4 结果：表5展示“EERIE平均尝试次数=5.2次”，误差必须用定量指标（如 $R^2$ 、MAE、RMSE）来支撑，不能只说“误差很小”，要给出具体数值；解读：“EERIE尝试次数高，反映‘重复字母+低词频’是高难度核心特征，符合用户反馈”；  5 对比：Stacking模型较单一ARIMA， $R^2$ 提升0.18。

目录章节	共性行文思路（固定逻辑）	O奖案例示例（以Wordle游戏建模赛题为例）
5.Sensitivity and Robustness Analysis	1. 敏感度分析：关键参数变化+结果影响+现实意义； 2. 稳健性分析：改数据/假设+结果波动+稳定性结论；	1. 敏感度分析（来自网球动量论文）对“破发成功率”这一关键参数进行敏感度分析：当破发成功率提升 10% 时，选手动量得分平均提升 15.3%。这一结果表明，破发是网球比赛中动量波动的核心驱动因素，现实中教练可通过强化破发战术来主动改变比赛走向。  2. 稳健性分析（来自奥运奖牌论文）放松“参赛国家数稳定”的初始假设后，重新输入波动 $\pm 10\%$ 的国家数量数据，模型预测结果的波动幅度 $\leq 3\%$ ，说明模型对短期参赛规模变化不敏感，稳健性良好，可适用于未来奥运奖牌预测场景。
6.Strength and Weakness	1.优势：用数据说话，突出「创新 + 精准 + 实用」。 创新：模型独特设计（如融合 / 改良）。精准：用 RMSE、R <sup>2</sup> 、F1 值等量化指标对比。 实用：说明落地价值（如为决策者提供参考）。 2.不足：客观聚焦「假设 / 数据 / 方法」局限。 假设：简化假设与现实的偏差。 数据：数据量 / 维度 / 可靠性不足。 方法：复杂场景下的适用性短板。 3.改进：针对不足给出具体可落地方案，形成闭环。	1.优势：用数据说话，突出「创新 + 精准 + 实用」。 创新：模型独特设计（如融合 / 改良）。精准：用 RMSE、R <sup>2</sup> 、F1 值等量化指标对比。 实用：说明落地价值（如为决策者提供参考）。 2.不足：客观聚焦「假设 / 数据 / 方法」局限。 假设：简化假设与现实的偏差。 数据：数据量 / 维度 / 可靠性不足。 方法：复杂场景下的适用性短板。 3.改进：针对不足给出具体可落地方案，形成闭环。
References	1. 数量：6-10篇； 2. 格式：统一APA或者ieee（选一个）； 3. 类型：领域经典+近年研究。	引用《Wordle: A Quantitative Analysis of Word Difficulty》(2023) 等6篇文献，格式统一。
Appendix	1. 代码：核心模型代码（含注释）； 2. 补充数据：完整得分表； 3. 公式推导：GMM聚类的详细步骤。 4.ai报告 5.letter或海报	1. 代码：ARIMA-LSTM融合代码（Python）； 2. 补充数据：2023年Wordle单词得分全表； 3. 公式推导：GMM后验概率的详细推导。

#### 4. 结尾升华（结论 + 价值 + 局限）

- 原创见解：基于模型结果给出落地建议（如给 IOC 的奥运建议、给教练的网球策略、给纽约时报的 Wordle 优化方案），不脱离赛题。
- 模型优劣：客观分析（如“GSRF 抗过拟合但未考虑变量交互”“SVM 对‘难词’分类较弱”），不回避局限。
- 格式规范：参考文献统一（APA或者ieee 格式，二选一），附录含代码 / 补充数据，图表有“标题 + 图例 + 编号”（如“Figure 1 2024 奥运奖牌榜”），公式编号连续。

### ▼ 八、建模逻辑共性与其他注意（高分关键）

#### 1. 问题驱动，不偏离赛题

- 所有模型、分析均围绕赛题任务，无冗余内容（如 Wordle 论文所有模型都服务于“预测结果分布、分类单词难度”，网球论文模型聚焦“动量量化、转向预测”）。
- 任务间衔接自然（如奥运论文 Task1 的 GSRF 结果直接作为 Task3 “项目 - 奖牌关系”的输入数据），形成逻辑闭环。

#### 2. 模型选择与创新（适配性 + 差异化）

- 拒绝盲目堆砌复杂模型：优先“经典模型改良”或“多模型融合”（如 ARIMA-LSTM、GSRF（网格搜索优化随机森林）、MIMO XGBoost），每个模型都说明“为何适配该任务”（如 LSTM 适配时间序列、Logistic 适配二分类）。
- 创新点明确且验证：或改良模型（如 GSRF 优化超参数），或提出新指标（动态竞赛指数 DPI、平均尝试次数 ATN），创新点必用定量指标验证（如“GSRF 比 ARIMA 更优，R<sup>2</sup> 从 0.494 提升至 0.784”）。

### 3. 数据支撑，拒绝空泛

- 每个结论都有“数据 + 公式”佐证：如“伟大教练对奖牌有显著影响”，用 Spearman 相关系数 0.874+Lasso 回归方程证明，而非单纯文字描述。
- 可视化辅助论证：用雷达图展示特征重要性、误差棒图展示预测区间、热力图展示相关性、混淆矩阵展示分类效果，让评审快速 get 核心结论。

1.对话中提到，M和O的主要区别在于数据和公式的支撑上。M可能更侧重于语言描述和概念解释，而不是严格的数学公式和数据支持。

2.如果英文水平不是非常高大家一般会先用中文写作，然后用翻译软件翻译再做修改，所以还要留出翻译和写作的时间。

#### 结果分析：让模型“落地”，体现价值

“算出结果”是基础，O 奖论文的核心竞争力是“结果解读的深度与实用性”（高奖与普通奖的分水岭）。

#### 结果呈现：量化为主，图表为辅

- 拒绝模糊描述，所有结论需数据支撑（例：“毛里塔尼亚获得首枚奖牌的概率 24.8%”“聘请优秀教练使获奖概率提升 3.74 倍”）；
- 小体量数据用表格，大体量用图表；每幅图 / 表需配 3 行以上分析，说明“数据反映的问题”与现实场景的关联。

#### 模型验证：多维检验，彰显严谨

O 奖论文必做“验证闭环”，核心含三项：

- 误差分析：用历史数据拟合、残差检验说明结果可信度（例：“模型预测值与历史数据误差控制在 5% 以内，拟合优度  $R^2=0.92$ ”）；
- 灵敏度分析：分析关键参数变化对结果的影响，解释现实意义（例：“除冰剂成本上涨 10%，最优分配方案向环境友好型产品倾斜 8%”）；
- 稳健性检验：改变假设或数据来源，验证模型稳定性（例：“放松‘人口流动稳定’假设后，预测结果波动幅度  $\leq 3\%$ ，模型稳健性良好”）。

### 路径1：AI生成框架 → Nanobanana精细化绘制 → 导出学术插图



### 路径2：现有简陋插图 → Nanobanana导入 → 美化优化 → 导出学术插图



☒ Thought · 1s

好了，现在运行程序：

☒ Run in terminal

```
cd /Users/xiaohuiwei/Download  
&& source .venv/bin/activate  
src/main.py
```

Terminal Output >

☒ Thought · 1s

有一个文件编码问题。让我修复它：

1. 文献综述和letter要写的；
2. 考试的时候可以补充数据吗 看情况， 扩大范围加州山火-南美山火。 a地光污染-b地光污染 可信度， 结论 看情况；
3. 科研团队导致常规模型隐患、是否用复杂（模型算法方面的），寻找更好建模的逻辑，减少回溯，看行业前沿；consensus 科研团队最新的结果复现（重要），权威团队，可复现性高；
4. 降ai率，风格上m： ai报告中要写：1.图制作；2.代码辅助debug
- 5.nano使用：可以gemini 生成提示词，再给nano 无底白图。部分图。拼在一起。文字加一加