**1、数据来源**

https://www.kaggle.com/datasets/amitvkulkarni/hair-health

**2、项目背景**

随着年龄增长，脱发成为许多人关注的健康问题之一。头发的丰盈与否不仅影响着外貌，更与个体的健康状态息息相关。

本数据集汇集了各种可能导致脱发的因素，包括遗传因素、荷尔蒙变化、医疗状况、药物治疗、营养缺乏、心理压力等。

通过数据探索分析，可以深入挖掘这些因素与脱发之间的潜在关联，从而为个体健康管理、医疗干预以及相关产业的发展提供有益参考。



**3、数据字典**

| **字段** | **说明** |
| --- | --- |
| Id | 标识符 |
| Genetics | 是否有秃头家族史（1:是 / 0:否） |
| Hormonal Changes | 是否经历了荷尔蒙变化 |
| Medical Conditions | 可能导致秃头的病史；斑秃、甲状腺问题、头皮感染、银屑病、皮炎等 |
| Medications & Treatments | 可能导致脱发的药物治疗史；化疗、心脏药物、抗抑郁药、类固醇等 |
| Nutritional Deficiencies | 营养不足情况 ；铁缺乏、维生素D缺乏、生物素缺乏、Ω-3脂肪酸缺乏等 |
| Stress | 压力水平 |
| Age | 年龄 |
| Poor Hair Care Habits | 是否存在不良的护发习惯 |
| Environmental Factors | 是否暴露于可能导致脱发的环境 |
| Smoking | 是否吸烟 |
| Weight Loss | 是否经历了显著的体重减轻 |
| Hair Loss | 是否脱发 |

**病史字段（Medical Conditions）**选项翻译：

| **英文** | **中文** |
| --- | --- |
| Alopecia Areata | 斑秃 |
| Androgenetic Alopecia | 雄激素型秃发 |
| Dermatitis | 皮炎 |
| Dermatosis | 皮肤病 |
| Eczema | 湿疹 |
| Psoriasis | 银屑病 |
| Ringworm | 癣 |
| Scalp Infection | 头皮感染 |
| Seborrheic Dermatitis | 皮脂溢出性皮炎 |
| Thyroid Problems | 甲状腺问题 |

**药物治疗史字段（Medications & Treatments）**选项翻译：

| **英文** | **中文** |
| --- | --- |
| Accutane | 治疗痤疮的药物 |
| Antibiotics | 抗生素 |
| Antidepressants | 抗抑郁药 |
| Antifungal Cream | 杀菌霜 |
| Blood Pressure Medication | 降血压药 |
| Chemotherapy | 化疗 |
| Heart Medication | 心脏药物 |
| Immunomodulators | 免疫调节剂 |
| Rogaine | 生发素 |
| Steroids | 类固醇 |

**营养不足字段（Nutritional Deficiencies）**选项翻译：

| **英文** | **中文** |
| --- | --- |
| Biotin Deficiency | 生物素缺乏 |
| Iron deficiency | 铁缺乏 |
| Magnesium deficiency | 镁缺乏 |
| Omega-3 fatty acids | Ω-3脂肪酸 |
| Protein deficiency | 蛋白质缺乏 |
| Selenium deficiency | 硒缺乏 |
| Vitamin A Deficiency | 维生素A缺乏 |
| Vitamin D Deficiency | 维生素D缺乏 |
| Vitamin E deficiency | 维生素E缺乏 |
| Zinc Deficiency | 锌缺乏 |

**压力水平字段（Stress）**选项翻译：

| **英文** | **中文** |
| --- | --- |
| Low | 低 |
| Moderate | 中 |
| High | 高 |

**数据来源**

<https://www.kaggle.com/datasets/amitvkulkarni/hair-health>

**4、可研究的问题**

**描述统计：**

1. 平均年龄是多少？年龄的分布情况如何？
2. 哪些医疗条件是最常见的？它们的发生频率是多少？
3. 营养缺乏的种类及其出现频率如何？

**可视化：**

1. 不同年龄段中患者脱发的比例是多少？可以通过柱状图或饼图展示。
2. 哪些因素与脱发之间存在关联？可以使用热力图或相关性图来呈现各因素之间的相关性。
3. 不同压力水平下脱发的情况如何？可以通过箱线图或分布图进行展示。

**机器学习：**

1. 可以建立一个分类模型来预测个体是否会患上脱发，以基于给定的因素进行预测。
2. 利用聚类分析，探索数据集中是否存在不同类型的脱发群体。
3. 使用决策树或随机森林等算法，识别出最能预测脱发的关键因素是什么。

**5、可参考的项目**

（1）**脱发👴的危险因素分析（Boruta+模型+SHAP流程）**

https://www.heywhale.com/mw/project/65fd94d847127ed3921e0c4b

（2）**基于统计检验与机器学习探索脱发的主要因素**

https://www.heywhale.com/mw/project/65d54a8ae33ab6bc1d5ed0a7

（3）**【每周挑战127期】脱发数据集的探索，关注聚类分析**

https://www.heywhale.com/mw/project/65d49706f231de650ff34667

（4）**关于脱发数据集的探索**

https://www.heywhale.com/mw/project/65d718dfade4eef0371ecb4d

（5）**【每周挑战127期】脱发数据集的建模学习**

https://www.heywhale.com/mw/project/65d35683e33ab6bc1d47890a

（6）**【🐳每周挑战】关于脱发数据集的探索**

https://www.heywhale.com/mw/project/65d312cae33ab6bc1d438143