1. C题 建立怎样的模型去解释自己的想法

* Address（解决） and discuss（讨论） ***whether*** or not the spread of this pest over time can be **predicted,** and with ***what*** level of precision.

地域因素：研究以红点为中心方圆8公里范围以内的区域，通过地理计算出经纬度，然后在这个经纬度范围内抽取报告

在这些范围以外的材料和陈述都视为无效提供

在地域范围内再对文本进行一个筛选，人工提取关键词，假设部分死亡的蜜蜂是由胡蜂造成，提取出报告时间，并假设一个死亡蜜蜂是由一个胡蜂造成，以2019~2020两年的月份（季度）为时间轴，胡蜂出现次数为纵坐标，标出点集。再利用matlab进行曲线拟合，通过图像分析，理论解释**the spread of this pest over time can be predicted**

给出解析式，再求出拟合优度R（要加上数学推导公式和拟合代码），完成***what* level of precision**。

然后是一个元胞自动机，黄蜂会向周围扩散，图像直接在地图上画出。

~~首先考虑空间差异性，利用Tablecu对给定的地图进行一个评分筛选（气温、湿度、食物、栖息环境，地形，经纬度），（假设其余条件相同）以华盛顿红色标记最多的地方为标准，选出图上适宜生存的地点，标记好经纬度的范围区域~~

* Most reported sightings mistake other hornets for the Vespa mandarinia. Use only the data set file provided, and (possibly) the image files provided, to ***create***, ***analyze***, and ***discuss*** a **model** that predicts the likelihood of a mistaken classification.

建立错误分类模型，将错认为胡蜂的情况进行一个分类(地域类，时间类，文本类，图像类)，（需要立即学习分类模型的建立）

假设华盛顿地理环境胡蜂宜居，计算每一个概率时要有理有据

1. 针对时间类：冬天几乎不可能出现胡蜂，百分百的错误
2. 针对地域类：凡是经纬度不在合适地理位置的地方的视为错误
3. 图像不清楚：把图像不清晰的视为错误概率大，因为不能识别
4. 解释不清楚：文本中对胡蜂的特征描述词太少或没有的错误概率大
5. 在前4条都满足的条件下，计算依然不为胡蜂的概率为多大，如果很小则模型可信。

最后将所有的情况分类，给出出现的案例和概率，错误可能性的概率，和真实出现错误的概率。

* ***Use*** your model to ***discuss*** **how** your classification analyses leads to prioritizing investigation of the reports（调查报告） most likely to be **positive sightings**.

利用二问中的概率模型，建立一个评分机制（层次分析法），文本中评分较高的图像优先进行及其图像分析和专业观察人员考察，不能清晰化的模糊照片视为unverified，清晰的照片分为positive和negative。评分低的优先度低。并用该模型对剩下的unprocessing的图片进行判断（用于后面report提建议）

* Address **how** you could ***update*** your model given additional new reports over time, and **how often** the updates should ***occur***.

假设胡蜂未得到控制，利用第一问的函数进行时间序列预测2021~2025年的情况（假设不改变其它外在环境条件），以及元胞自动机对空间进行预测，随着时间的推移，地域范围会扩大，所得到的positive的图像数据也逐渐全面，机器可以进行学习，同时地域的扩大会使得部分地区错误概率会增加。

考虑胡蜂生活习性，夏天更新一次，春秋冬合起来更新一次。（地域和图像）

考虑害虫防止越早越好，所以一年两次比较合理

* Using your model, **what** would constitute evidence that the pest has been eradicated（根除） in Washington State?

当胡蜂出现的数量减少到认定范围内，就可以判断已经根除（查阅文献）

* Finally, your report should include a two-page **memorandum**（备忘录） that **summarizes** your results for the Washington State Department of Agriculture.

给政府写一个两页备忘录，包括概述背景，解释模型如何分类错误，以及怎样的打分机制给出优先度，演示怎么对unprogressing的材料判断以提高政府工作效率。还有对未来的时间和空间上胡蜂的预测，更新机制的具体时间，以及给政府提供一些治理措施。 follow-up with additional investigation

最后加上灵敏度分析和模型优缺点。

文献每次都要记录，改进的模型，参考的论文都要有标注

论文写作，在谷歌翻译后要人工检查，表达精确无误

要做到图像可视化，不光是数据图像，也要有小蜜蜂的照片

论文要写的有趣和有逻辑，排版要工整。