**第一问**

Address and discuss whether or not the spread of this pest over time can be predicted, and with what level of precision. 解决并讨论是否可以预测这种害虫随时间的传播，以及以何种程度的精度。（精确到何种程度）

预测方法：先通过情感分析进行notes的得分，再通过对各列进行打分，增加正例个数，最后通过拟合函数就行拟合预测。

**第二问**

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. 大多数报道称目击到的黄蜂都把其他黄蜂误认为了大黄蜂。仅使用所提供的数据集文件和(可能)所提供的图像文件来创建、分析和讨论预测错误分类可能性的模型。

首先需要通过图像识别将所有图片、视频中的昆虫找到，然后通过与第一问得到的正例进行匹配，然后通过模型训练。

图像识别模型

<https://blog.csdn.net/qq_19582165/article/details/102439503?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522164413059816780269845425%2522%252C%2522scm%2522%253A%252220140713.130102334..%2522%257D&request_id=164413059816780269845425&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduend~default-2-102439503.pc_search_insert_ulrmf&utm_term=%E5%9B%BE%E5%83%8F%E8%AF%86%E5%88%AB%E6%A8%A1%E5%9E%8B&spm=1018.2226.3001.4187>

训练模型：https://blog.csdn.net/qq\_42370150/article/details/103084897

**第三问**

Use your model to discuss how your classification analyses leads to prioritizing investigation of the reports most likely to be positive sightings. 使用您的模型来讨论您的分类分析如何导致优先调查最有可能是正面目击的报告。

阐述模型可行性。

**第四问**

Address how you could update your model given additional new reports over time, and how often the updates should occur. 说明如何随着时间的推移对额外的新报告更新您的模型，以及更新应该发生的频率。

maybe随着时间的推移，第二问中的一些参数权重发生变化？

或者机器学习自学习？

**第五问**

Using your model, what would constitute evidence that the pest has been eradicated in Washington State? 用你的模型，有什么证据能证明华盛顿州已经根除了这种害虫?