**4. NLP Task**

OCEAN scores analysis: Suppose you’re given the task to build a tool that takes a user’s tweet data (all the tweets they’ve ever made), and predicts their OCEAN scores to an 80% accuracy. Let’s assume we can get OCEAN scores against 100k twitter users.

This is a lofty goal to achieve even with this size of data, and we must plan and tackle the problem smartly.

**1.** The first logical step is to find research papers that have done something similar to set a benchmark. How will you find papers against this problem? List the most promising papers you find.

**2.** How will you plan you experimentation towards building a robust model? What techniques can you employ based on your initial research?

**3.** How will you document your efforts and report progress?