Team's name:
ChronicSaviours

Team members:
Seyedhooman Sajjadi
Anvay Govind Pandit
Shibin Tazhe Vettil

- What is the need? Who wants or benefits?

Many people experience suffering chronic diseases and symptoms throughout their life. Chronic symptoms usually need persistent management and keeping track of, since there is usually no certain cure for them. Patients usually go through a long period of treatment with different medications and remedies while tracking the effect.
Analysing and interpreting the relationships between symptoms, treatments, and the environmental factors to find meaningful patterns can help doctors better understand the nature of these diseases and better help the patients managing the symptoms.

- What data (or datasets)?

We are going to use Flaredown **Chronic illness: symptoms, treatments and triggers[1]** database. This dataset was mainly created to investigate how treatments and environmental stressors impact patient's symptoms.

The users were asked to create their own unique set of conditions, symptoms, and treatments so called "trackables". There are around 2 millions rows each for a single day, consisting of the patient's ID, age, sex, and country, the patient's condition, the symptoms as well as their severity, the dose of treatments, food, weather, Harvey Bradshaw Index(HBI), and that day's unexpected environmental factors called 'tags'.

- What is your "data mining" toolkit? You should list specific methods you will implement.

Pre-Processing and Data Cleaning - Some users who do not see a predefined symptom,treatment,tag,or food in the database may simply add it by giving a named description to these fields. So we need to clean such rows of data using regular expressions and extracting common terms out of them; and then map them to our predefined values. We might also need to do topic modelling using LDA if the handcrafted values are not enough.

To guess a condition based on user's symptoms we plan to use decision trees and SVM's by treating the conditions as classes. We will report the classifier giving better results.

To recommend users with effective treatments, we plan to find similar users(user-user similarity) and recommend their treatments. To find similar users we plan to use KNN Similarity/SVD/SVD++.

Data Visualization with Peak detection : To Analyze the effect of a treatment on a set of systems across users, we can make use of the timestamps and detect a peak on the severity index to analyze the effectiveness of a treatment.We can plot an example process using visualization tools like Tableau/Seabourn.

- Preliminary sketch of what you hope to find.

We primarily have the following objectives:

1) To find Correlations between symptoms and treatments.
2) To find if a treatment was effective or not. (Peak detection)
3) To guess the chronic condition based on user's symptoms(SVM/Decision Tree Classifier)
4) To recommend treatments based on similarity of user's rather than specific symptoms. (User-User Similarity)

References:

[1] https://www.kaggle.com/flaredown/flaredown-autoimmune-symptom-tracker