

TEAM MEMBERS DETAILS

GROUP NAME: Health Research

Name	Email	Country	College/Company	Specialisation
Kelvin Mpofu	mpofukelvintafadzwa@gmail.com	South Africa	n/a	Data science
Purity Nyagweth	purityeverniter@gmail.com	Kenya	n/a	Data Science
Reshma Jayapalan	reshma.jayapalan@gmail.com	UAE	n/a	Data Science
Hanouf Hazza	hanouf.haz@gmail.com	Saudi Arabia	n/a	Data Science

Problem Description

ABC pharma company has a challenge in understanding drug persistency as per physician prescription and to solve this problem it wants to automate the process of identification.

Objective is to build a classification model for drug persistency identification.

This will automate the process of identifying drug persistency for ABC pharma company thus helping the company to understand drug persistency as per physician prescription.

Business understanding

Understanding the persistency is an issue for pharmaceutical companies and so, ABC pharma company has approached XYZ analytics company to provide insights into the same.

The role of XYZ analytics company is to undertake this project and provide a detailed understanding regarding the drug persistency.

Projected Business advantages:

- Adherence to a drug directly impacts the sales dollars a product can generate
- Optimize and improve the efficacy of clinical trials
- Target specific patient populations more effectively
- Better insight into patient behaviour to improve drug effectiveness

Extra details

When patients are prescribed treatment by a health care official, it is expected that the patient adheres to the prescribed treatment regimen. Medication persistence refers to the act of continuing the treatment for the prescribed duration. It may be defined as the duration of time from initiation to discontinuation of therapy [1]. Inadequate medication persistence is an age old problem which can have situation-specific alterations in benefit/risk ratios, either because of reduced benefits, increased risks, or both [1].

Though not of interest in this particular case it should be known that Medical compliance is also a problem. Medication compliance (also known as adherence) refers to the degree or extent of conformity to the recommendations about day-to-day treatment by the provider with respect to the

timing, dosage, and frequency. It is described as the extent to which a patient acts in accordance with the prescribed interval, and dose of a dosing regimen [1].

Studies have shown that inadequate compliance and non-persistence with prescribed medication regimens result in increased morbidity and mortality from a wide variety of illnesses, as well as increased health-care costs in particular cases [2,3]. Clinical outcomes of treatment depend not only by how well patients take their medications but also by how long they take their medications.

There are many factors which affect a patients persistence such as choice of agent prescribed, comorbidity, and socioeconomic status, despite universal coverage of prescription drug costs [3]. It is important for the sake of improving clinical outcomes to be able to know which patients will less likely to stick with the medication persistence requirements and which ones will not.

Machine learning algorithms are particularly useful for such problems [4]. Our objective is to apply machine learning to predict patient medical persistence with the aim of using this information to identify patients who are least likely to adhere to medical persistence [4]. This information may be useful in helping patients.

Nontuberculous mycobacterial (NTM) lung disease is a general term which refers to a group of disorders characterized by exposure to specific bacterial germs known as mycobacteria [5]. These germs are found in the water and soil and are common throughout the environment as a whole. They usually do not cause illness. In NTM disorders, severity of infection and the disease course can vary greatly from one person to another. The most common symptoms include a persistent cough, fatigue, weight loss, night sweats, and occasionally shortness of breath (dyspnea) and coughing up of blood (hemoptysis) [5].

We will examine a dataset of patients who have NTM. There is a wide range of features which we will use to predict medical persistence in patients. These include but are not limited to patient demographics, provider attributes, clinical factors and disease factors.

[1] J. Cramer, A. Roy, A. Burrell, C. Fairchild, M. Fuldeore, D. Ollendorf and P. Wong, Medication compliance and persistence: Terminology and definitions, *Value in health*, 2, 1(2008).

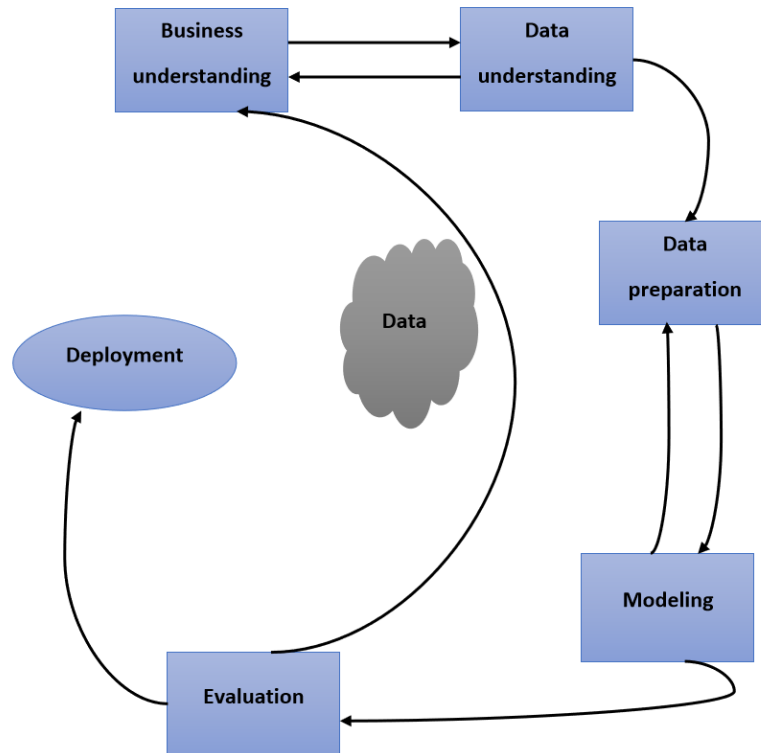
[2] M. DiMatteo, P. Giordani, H. Lepper and T. Croghan, Patient adherence and medical treatment outcomes: a meta-analysis, *Med Care*, 40, 794–811 (2002).

[3] J. Avorn, J. Monette, A. Lacour A, R. Bohn, M. Monane, H. Mogun and J. Ilerier, Persistence of use of lipid-lowering medications: cross-national study, *JAMA*, 279, 1458–62 (1998).

[4] X. Wu, H. Yang, R. Yaun, E. Long and R. Tong, Predictive models of medication non-adherence risks of patients with T2D based on multiple machine learning algorithms, *BJM Open Diabetes and Care*, 8, 1 (2020).

[5] <https://rarediseases.org/rare-diseases/nontuberculous-mycobacterial-lung-disease/>

Project lifecycle



The schedule below clarifies the project lifecycle, tasks, and deadlines. The deadlines are split into two, the first draft will be every Friday, and the official one would be on Sundays. Using Github for work and tasks sharings.

Phases	Tasks	Deadlines
Ph1. Understanding the problem	<ul style="list-style-type: none"> - Data understanding - Business understanding 	Week7 1 st draft (09/03/2021) Official (09/05/2021)
Ph2. Data preparation	<ul style="list-style-type: none"> - Data cleaning - Type of data - Data problems - Approaches used to overcome the problems in the data - Techniques used 	Weeks 8 1 st draft (09/10/2021) Official (09/12/2021) Week9 1 st draft (09/17/2021) Official (09/19/2021)
Ph3. Model planning	<ul style="list-style-type: none"> - Exploratory Data Analysis (EDA) - EDA performed on data - EDA recommendation 	Week10 1 st draft (09/24/2021) Official (09/26/2021)

	- EDA presentation for business users	
Ph4. Model building	- Linear model - Ensemble - Boosting	Week11 1 st draft (09/01/2021) Official (09/03/2021)
Ph5. Communicate results	- Findings are shared with the stakeholders - Ppt presentation	Week12 1 st draft (10/08/2021) Official (10/10/2021)

DATA INTAKE REPORT

Name: Data Science – Healthcare project

Report date: 05/09/2021

Internship Batch: LISUM02

Version:1.0

Data intake by: Health research

Data intake reviewer:<All group members>

Data storage location: github

Tabular data details:

Total number of observations	3424
Total number of files	1
Total number of features	69
Base format of the file	.xlsx
Size of the data	899 KB