

TECHNICAL REPORTS – RECOMMENDATION SYSTEM

Oleh: Christiani Turnip (2206130694)

Komputasi Lanjut dan Big Data

Dosen: Dr. Risman Adnan, S.Si., M.Si

Pendahuluan

Dalam bidang perawatan kesehatan, penulisan resep obat adalah proses yang kompleks dan mempengaruhi hasil pengobatan pasien secara signifikan. Pemilihan obat yang tepat dengan mempertimbangkan efektivitas, profil keamanan, dan karakteristik pasien menjadi tantangan bagi para profesional kesehatan. Kesalahan dalam penulisan resep obat dapat memiliki konsekuensi serius, seperti reaksi yang merugikan, pengobatan yang tidak efektif, dan bahkan dapat membahayakan nyawa pasien. Untuk mengatasi masalah ini, sistem rekomendasi obat telah muncul sebagai solusi yang menjanjikan.

Sistem rekomendasi obat menggunakan teknologi dan pendekatan berbasis data untuk membantu para profesional kesehatan dalam membuat keputusan penulisan resep yang terinformasi dan personal. Sistem ini menggabungkan informasi dari berbagai sumber, termasuk data pasien, pedoman klinis, basis data obat, dan pengetahuan medis yang relevan. Dengan menganalisis data tersebut, sistem rekomendasi obat dapat memberikan rekomendasi yang disesuaikan dengan kebutuhan dan karakteristik individu pasien.

Tujuan utama dari sistem rekomendasi obat adalah untuk meningkatkan kualitas pengobatan dan keselamatan pasien. Dengan menyediakan rekomendasi yang didasarkan pada bukti medis dan faktor-faktor individual pasien, sistem ini membantu mengurangi risiko kesalahan penulisan resep dan memastikan bahwa pasien menerima pengobatan yang paling sesuai untuk kondisi mereka.

Selain itu, sistem rekomendasi obat juga dapat membantu mengoptimalkan penggunaan sumber daya kesehatan secara efisien. Dengan menyediakan rekomendasi yang tepat, sistem ini dapat membantu mengurangi biaya perawatan kesehatan dengan mencegah penggunaan obat yang tidak diperlukan atau tidak efektif.

Namun, meskipun sistem rekomendasi obat menjanjikan, ada beberapa tantangan yang perlu diatasi. Salah satunya adalah masalah privasi dan keamanan data pasien. Penggunaan data medis sensitif dalam sistem ini memerlukan perlindungan yang kuat untuk memastikan kerahasiaan dan integritas informasi pasien.

Latar belakang

Penulisan resep obat merupakan salah satu aspek penting dalam praktik medis. Dokter harus mempertimbangkan banyak faktor, termasuk diagnosis pasien, karakteristik individu, riwayat medis, interaksi obat, dan pedoman klinis saat memilih obat yang paling sesuai. Namun, dengan berkembangnya pengetahuan medis dan peningkatan jumlah obat yang tersedia di pasaran, tugas ini semakin rumit dan menantang.

Kemajuan dalam teknologi dan bidang analisis data telah membuka peluang untuk mengembangkan sistem rekomendasi obat yang dapat membantu para dokter dalam pengambilan keputusan penulisan resep yang lebih terinformasi. Sistem rekomendasi obat menggabungkan pendekatan berbasis data dengan kecerdasan buatan untuk menganalisis data pasien, penelitian medis terkini, dan informasi obat untuk memberikan rekomendasi obat yang personal dan efektif.

Latar belakang penelitian dalam pengembangan sistem rekomendasi obat melibatkan peningkatan penggunaan data kesehatan elektronik (EHR) dan perkembangan teknik analisis data seperti pembelajaran mesin (machine learning) dan pemrosesan bahasa alami (natural language processing). Dengan menganalisis besar data pasien dan informasi medis terkait, sistem ini dapat mengidentifikasi pola, asosiasi, dan interaksi obat yang kompleks yang sulit untuk ditangani oleh manusia secara manual.

Sistem rekomendasi obat juga memanfaatkan basis data obat yang luas yang mencakup informasi tentang efek samping, interaksi obat, dosis yang dianjurkan, dan profil keamanan. Dengan menggabungkan data ini dengan informasi pasien dan pedoman klinis, sistem rekomendasi obat dapat memberikan rekomendasi yang personal, mempertimbangkan faktor seperti alergi, kondisi medis yang ada, riwayat pengobatan, dan karakteristik individu lainnya.

Dengan terus berkembangnya teknologi dan penelitian di bidang ini, diharapkan sistem rekomendasi obat akan menjadi alat yang berharga dalam praktik medis yang dapat meningkatkan kualitas pengobatan, meminimalkan risiko, dan mengoptimalkan hasil bagi pasien.

Data Collection

```
import numpy as np
import pandas as pd
import os
import spacy
import en_core_web_sm
from spacy.lang.en import English
from spacy.lang.en.stop_words import STOP_WORDS
import string
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
from sklearn.feature_extraction import DictVectorizer
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer, HashingVectorizer
from sklearn.base import TransformerMixin
from sklearn.pipeline import Pipeline
!pip install vaderSentiment
import vaderSentiment
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
```

```

!pip install -U -q PyDrive
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials
auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)

from google.colab import drive
drive.mount('/content/gdrive')

```

```

# preprocessing the data file
# read the data
df1 = pd.read_csv("/content/gdrive/MyDrive/Colab Notebooks/drugsComTrain_raw.csv")
df2 = pd.read_csv("/content/gdrive/MyDrive/Colab Notebooks/drugsComTest_raw.csv")
# combine two file
df = pd.concat([df1, df2])
df
# rename the cols
df.columns = ['ID', 'drug name', 'condition', 'review', 'rating', 'date', 'useful count']

```

Impor kumpulan data ke dalam proyek menggunakan modul baca di Pandas. Kode di atas telah dibagi menjadi train dan test set. Kemudian, kedua dataset digabungkan menggunakan Fungsi pandas concat, dan nama kolom dari kumpulan data yang digabungkan diubah.

Impor perpustakaan seaborn dan matplotlib untuk memvisualisasikan data. Tetapkan parameter terlebih dahulu, lalu hitung berapa rating yang memiliki nilai 10. Selanjutnya, buat diagram batang dengan sumbu x sebagai drugname dan sumbu y sebagai drug_rating.

```

import seaborn as sns
#!pip install matplotlib
import matplotlib.pyplot as plt

# Setting the Parameter
sns.set(font_scale = 1.2, style = 'darkgrid')
plt.rcParams['figure.figsize'] = [15, 8]

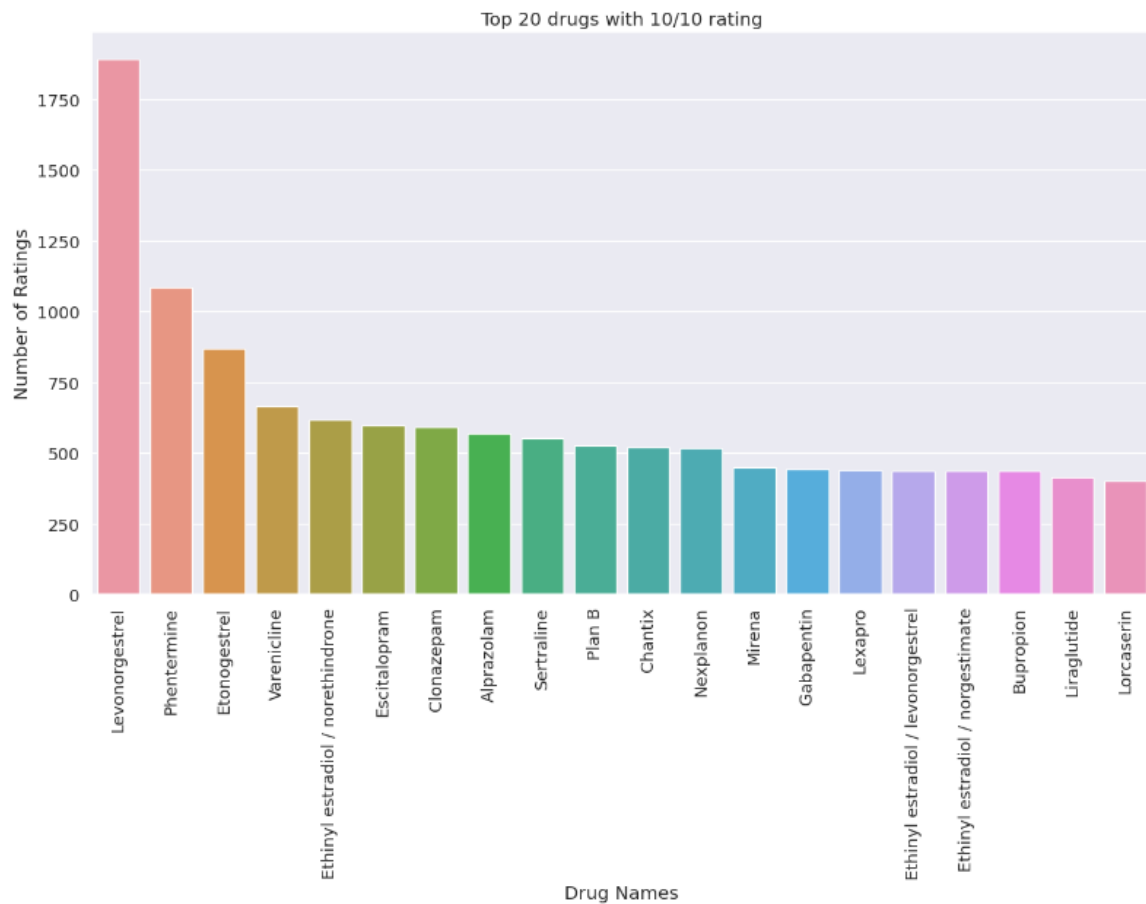
rating = dict(df.loc[df.rating == 10, "drug name"].value_counts())
drugname = list(rating.keys())
drug_rating = list(rating.values())

sns_rating = sns.barplot(x = drugname[0:20], y = drug_rating[0:20])

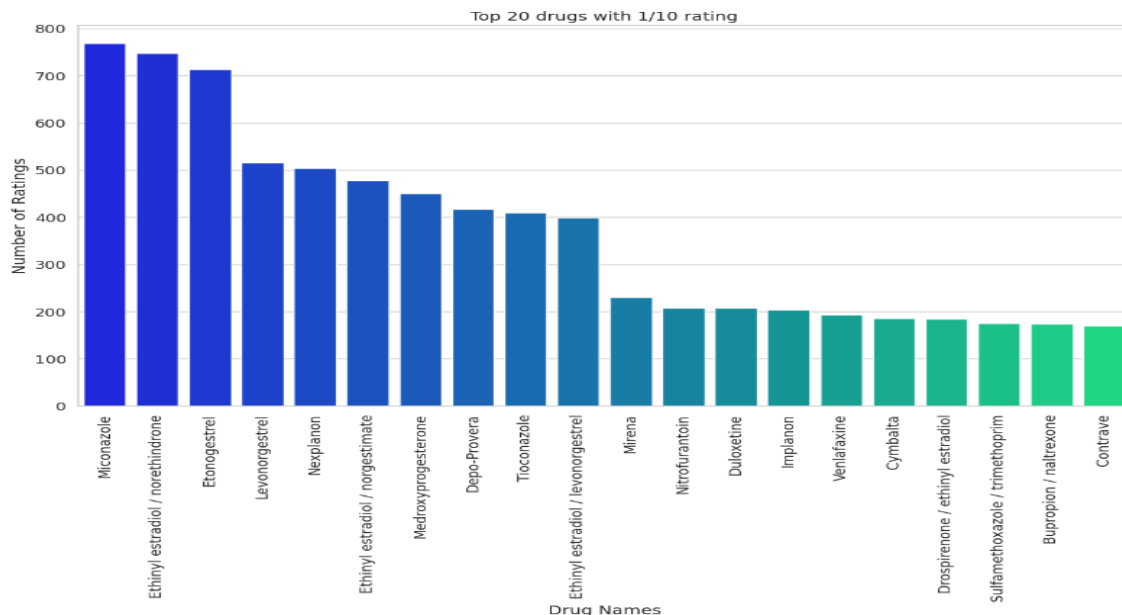
sns_rating.set_title('Top 20 drugs with 10/10 rating')
sns_rating.set_ylabel("Number of Ratings")
sns_rating.set_xlabel("Drug Names")
plt.setp(sns_rating.get_xticklabels(), rotation=90);

```

Output:



Pada output dapat dilihat bahwa ada 20 obat teratas total rating tertinggi 10. Diantaranya, obat Levonorgestrel memiliki total nilai rating tertinggi lebih dari 1750, diikuti oleh obat Phentermine dan lain-lain.



Dapat dilihat bahwa ada 20 obat teratas total rating tertinggi 1. Diantaranya, obat Miconazole memiliki total nilai rating tertinggi lebih dari 750, diikuti oleh obat Ethinyl estradiol dan lain-lain.

```

# A countplot of the ratings so we can see the distribution of the ratings
plt.rcParams['figure.figsize'] = [20,8]
sns.set(font_scale = 1.4, style = 'whitegrid')
fig, ax = plt.subplots(1, 2)
sns_1 = sns.countplot(x='rating', data=df, palette='spring', order=list(range(10, 0, -1)), ax=ax[0])
sns_2 = sns.distplot(df['rating'], ax = ax[1])
sns_1.set_title('Count of Ratings')
sns_1.set_xlabel("Rating")

sns_2.set_title('Distribution of Ratings')
sns_2.set_xlabel("Rating")

```

Output:

