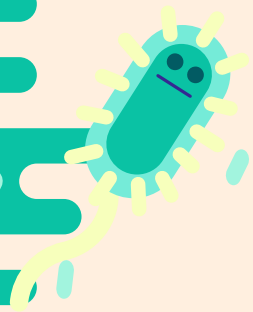
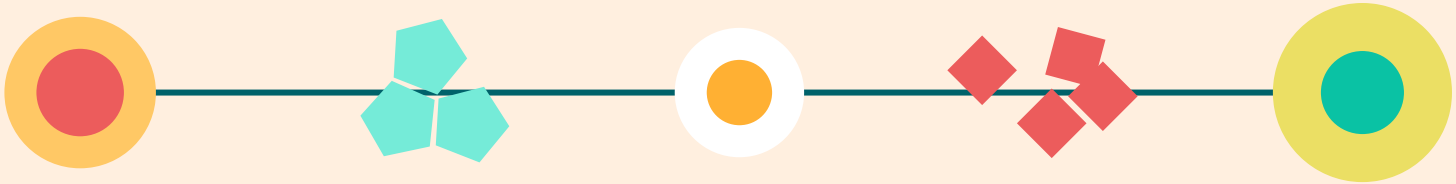
CYTOCHROME P450 INHIBITION

A capstone
&
A contribution to open source project: adme-pred-py



5 Enzymes

CYP2c19 CYP2c9 CYP1a2 CYP2d6 CYP3a4



The 5 Cytochrome P450 block of enzymes we are talking about today are a part of a larger block of more than 50 different enzymes.

These 5 are responsible for breaking down about 90% of drugs.

Genetics affect the way they metabolize proteins.

Because of this, there are HUGE implications for prescribing drugs that are ineffective or possibly toxic for the individual patient.

Current Treatment Experience

OUTDATED + INACCURATE

The majority of diagnostics science, and treatment are based on relatively healthy white males

GENE EXPRESSION

Genes expression for drug metabolism can vary which can have a dramatic effect on how a person uptakes and expels the drug

DANGERS

Treating ‘blind’ can result in side effect variance, variable treatment success, toxicity and even death.

GOALS

01

Tailored Medicine

Create tailored treatment plans based on cytochrome isoenzyme activity

02

Efficiency + Efficacy

With machine learning, medicine can be quickly tailored to the patient

03

Pharmacokinetics

These models can also aide in creating new drugs that utilize these specific enzymes.

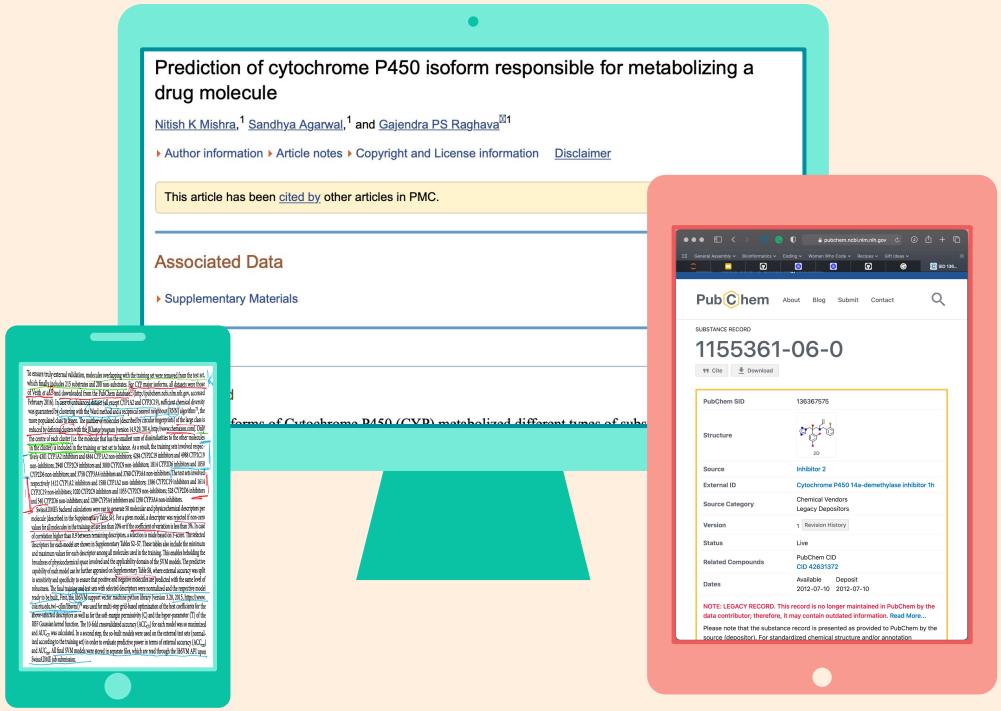
04

Patient Outcome

Overall patient outcome will be improved, with timeliness, and more effective treatment.

Research Process

- At least 20 research papers
- Pubchem
- Chembl
- Rdkit
- NCBI & Nature.com



adme-pred-py

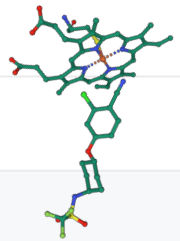
4 Open 7 Closed Author Label Projects Milestones Assignee Sort

Add basic testing
#12 opened 3 days ago by ikmckenz

Change main run function to accept input from terminal enhancement good first issue hacktoberfest 1
#10 opened on Sep 1, 2020 by ikmckenz

Add Cytochrome P450 inhibition ML model core-adme enhancement 5
#8 opened on Aug 28, 2020 by ikmckenz

Add Martin 'A Bioavailability Score' core-adme enhancement hacktoberfest
#6 opened on Mar 14, 2020 by ikmckenz



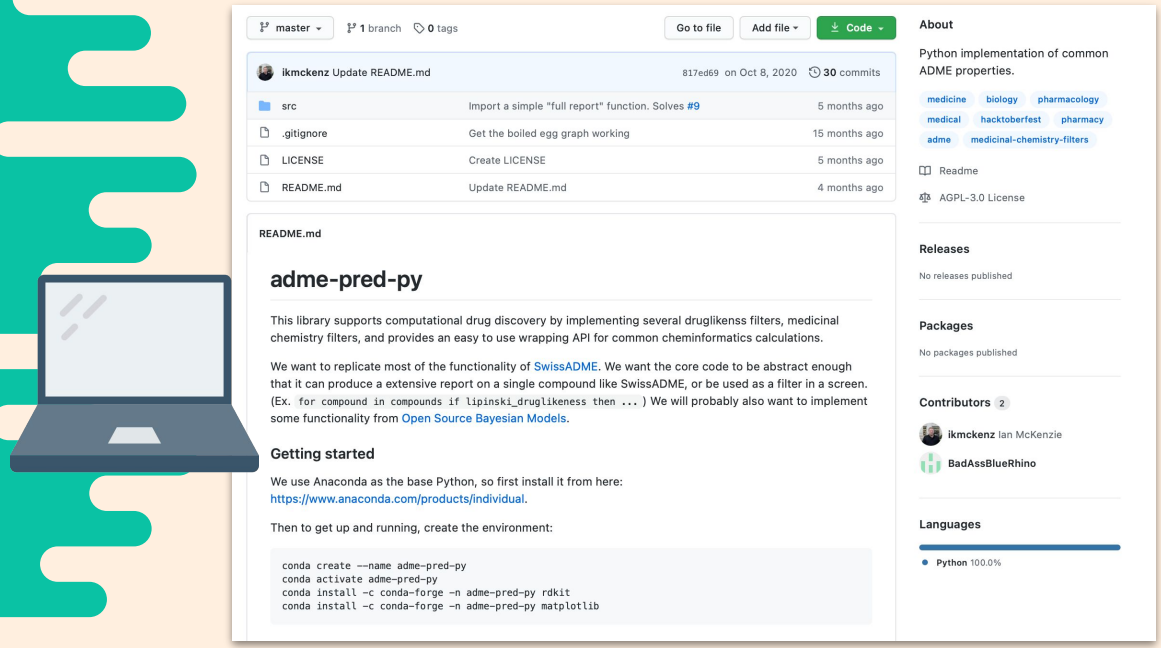
ProTip! What's not been updated in a month: updated:<2021-01-19.

CYP2c9

I stumbled upon an open source project that had an open issue specifically to create CYP450 inhibition machine learning model!

adme-pred-py

imkcenz/



master · 1 branch · 0 tags — Go to file · Add file · Code

ikmckenz Update README.md — 817ed69 on Oct 8, 2020 — 30 commits

src	Import a simple "full report" function. Solves #9	5 months ago
.gitignore	Get the boiled egg graph working	15 months ago
LICENSE	Create LICENSE	5 months ago
README.md	Update README.md	4 months ago

README.md

adme-pred-py

This library supports computational drug discovery by implementing several druglikenss filters, medicinal chemistry filters, and provides an easy to use wrapping API for common cheminformatics calculations.

We want to replicate most of the functionality of SwissADME. We want the core code to be abstract enough that it can produce a extensive report on a single compound like SwissADME, or be used as a filter in a screen. (Ex. `for compound in compounds if lipinski_druglikeness then ...`) We will probably also want to implement some functionality from Open Source Bayesian Models.

Getting started

We use Anaconda as the base Python, so first install it from here:
https://www.anaconda.com/products/individual.

Then to get up and running, create the environment:

```
conda create --name adme-pred-py
conda activate adme-pred-py
conda install -c conda-forge -n adme-pred-py rdkit
conda install -c conda-forge -n adme-pred-py matplotlib
```

About

Python implementation of common ADME properties.

medicine · biology · pharmacology · medical · hacktoberfest · pharmacy · adme · medicinal-chemistry-filters

Readme

AGPL-3.0 License

Releases

No releases published

Packages

No packages published

Contributors 2

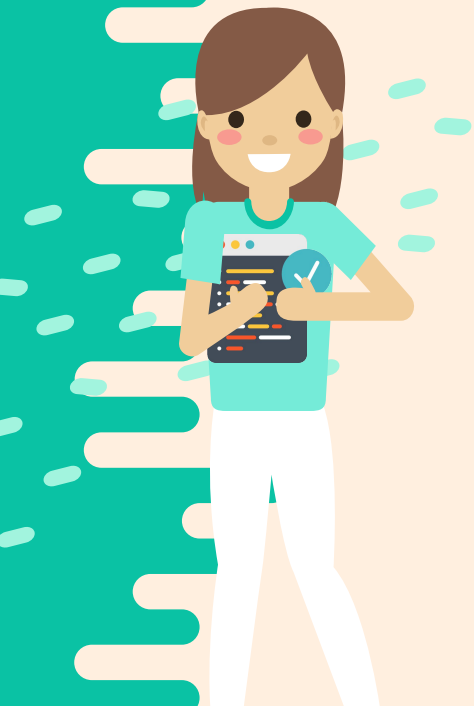
ikmckenz Ian McKenzie

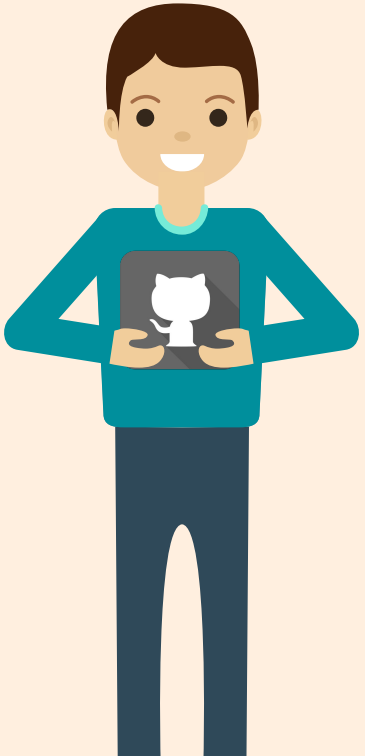
BadAssBlueRhino

Languages

Python 100.0%

Open Source Collaboration

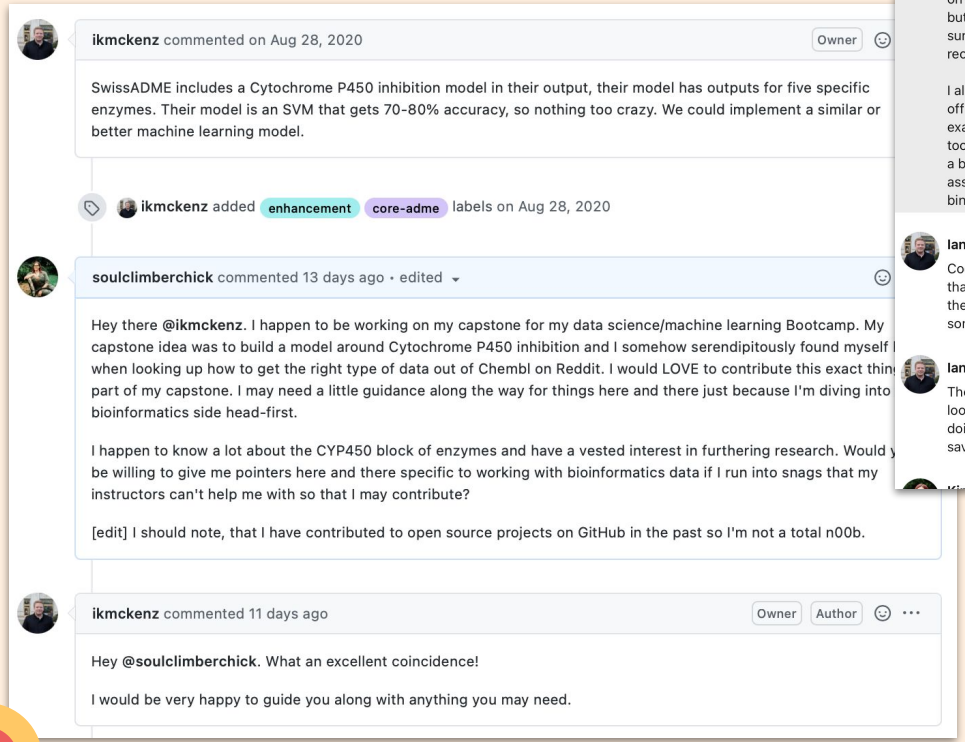




I reached out to the creator of the project.

Shared info about my capstone and asked if I could contribute to his project

Collaboration



ikmckenz commented on Aug 28, 2020 — Owner

SwissADME includes a Cytochrome P450 inhibition model in their output, their model has outputs for five specific enzymes. Their model is an SVM that gets 70-80% accuracy, so nothing too crazy. We could implement a similar or better machine learning model.

ikmckenz added enhancement core-adme labels on Aug 28, 2020

soulclimberchick commented 13 days ago • edited

Hey there **@ikmckenz**. I happen to be working on my capstone for my data science/machine learning Bootcamp. My capstone idea was to build a model around Cytochrome P450 inhibition and I somehow serendipitously found myself
when looking up how to get the right type of data out of Chembl on Reddit. I would LOVE to contribute this exact thin
part of my capstone. I may need a little guidance along the way for things here and there just because I'm diving into
bioinformatics side head-first.

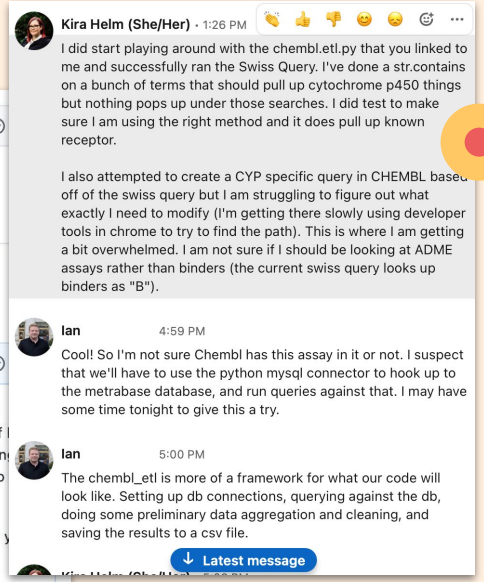
I happen to know a lot about the CYP450 block of enzymes and have a vested interest in furthering research. Would y
be willing to give me pointers here and there specific to working with bioinformatics data if I run into snags that my instructors can't help me with so that I may contribute?

[edit] I should note, that I have contributed to open source projects on GitHub in the past so I'm not a total n00b.

ikmckenz commented 11 days ago — Owner Author

Hey **@soulclimberchick**. What an excellent coincidence!

I would be very happy to guide you along with anything you may need.



Kira Helm (She/Her) • 1:26 PM

I did start playing around with the chembl.etl.py that you linked to me and successfully ran the Swiss Query. I've done a str.contains on a bunch of terms that should pull up cytochrome p450 things but nothing pops up under those searches. I did test to make sure I am using the right method and it does pull up known receptor.

I also attempted to create a CYP specific query in CHEMBL based off of the swiss query but I am struggling to figure out what exactly I need to modify (I'm getting there slowly using developer tools in chrome to try to find the path). This is where I am getting a bit overwhelmed. I am not sure if I should be looking at ADME assays rather than binders (the current swiss query looks up binders as "B").

Ian 4:59 PM

Cool! So I'm not sure Chembl has this assay in it or not. I suspect that we'll have to use the python mysql connector to hook up to the metrabase database, and run queries against that. I may have some time tonight to give this a try.

Ian 5:00 PM

The chembl_etl is more of a framework for what our code will look like. Setting up db connections, querying against the db, doing some preliminary data aggregation and cleaning, and saving the results to a csv file.

Latest message

Soulclimberchick > < Imckenz

Data Detangling

Feature Engineering

Pubchem ID

842319

Canonical SMILES

CC1=CC(=NO1)C(=O)NN=CC2=CC=CC=C2Br

Hashed Morgan Fingerprint

[1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, ...]
64 Bits

Morgan Bitstring

1100001011110010110101111000100111001011100101...

String

1100001011110010110101111000100111001011100101...

Features

|1 |2 |3 |4 |5 |6 |7...
|1 |1 |0 |0 |0 |0 |0....

Feature Creation Process

Enter: Obnoxiously long lambda function. (no this is not good practice but... time).

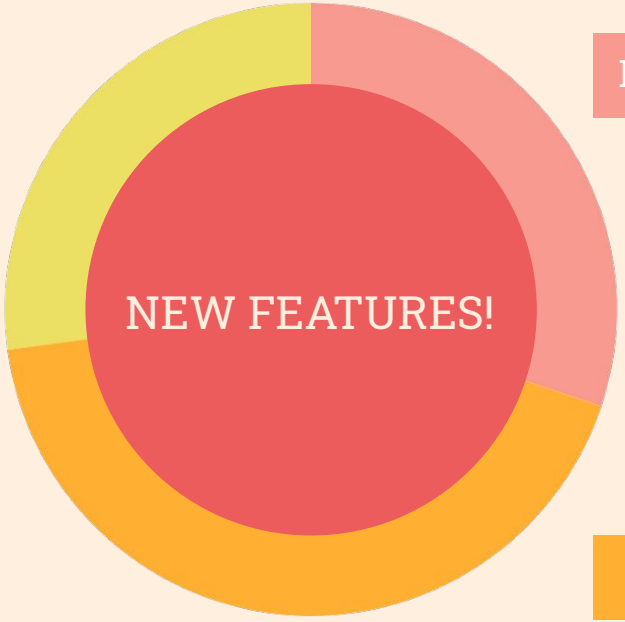
```
cyp2c19_512['MORGAN_BTSTR'] =

cyp2c19_512.apply(lambda row: AllChem.GetMorganFingerprintAsBitVect

(Chem.MolFromSmiles(row['SMILES']), 2, nBits=512)

.ToBitString(), axis=1)
```

But Wait, there's MORE!



Engineered Featuers

Rdkit Features

Fingerprint

SwissAdme Features

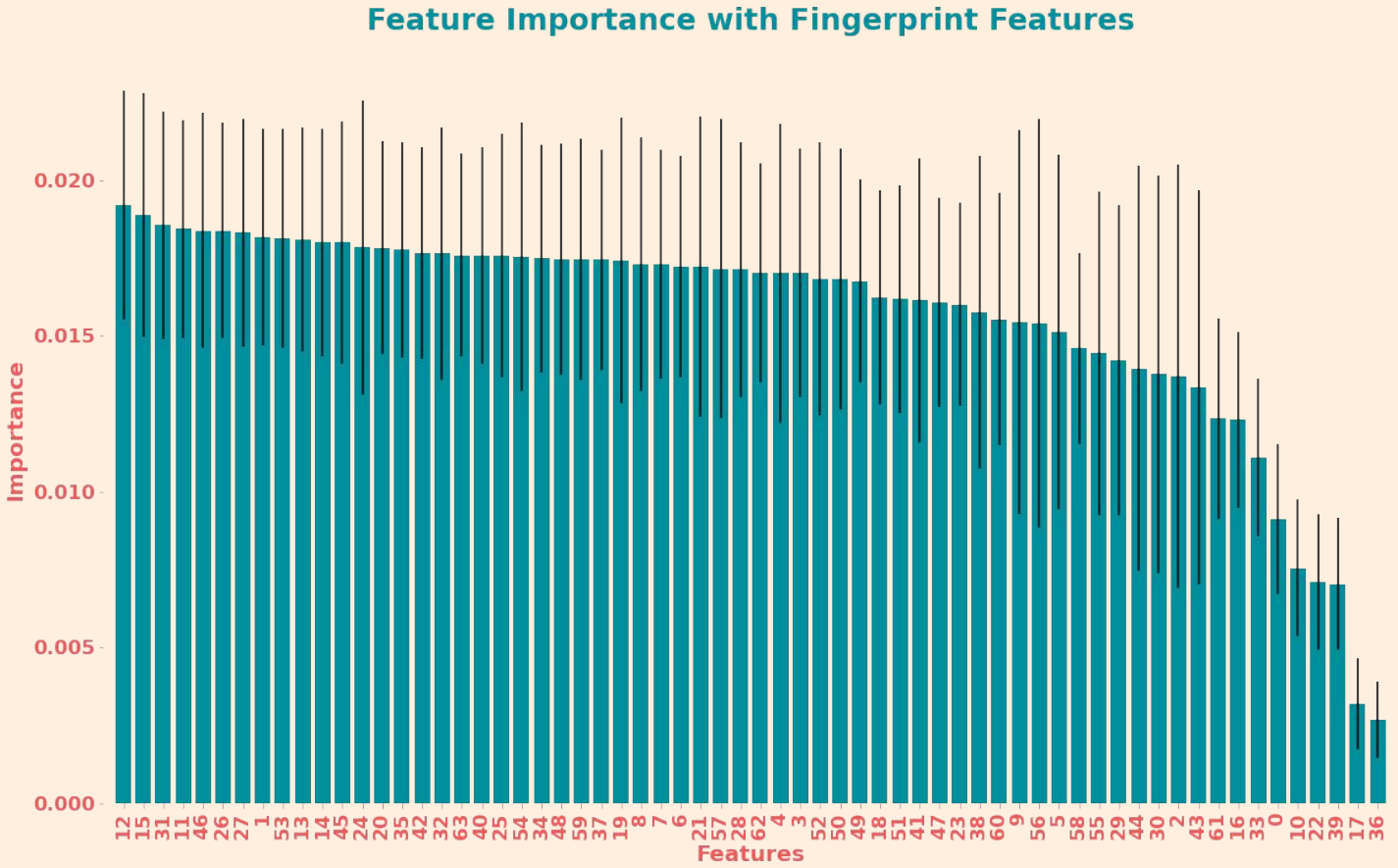
After working with fingerprint data, I decided to tackle using SwissAdme's feature set.

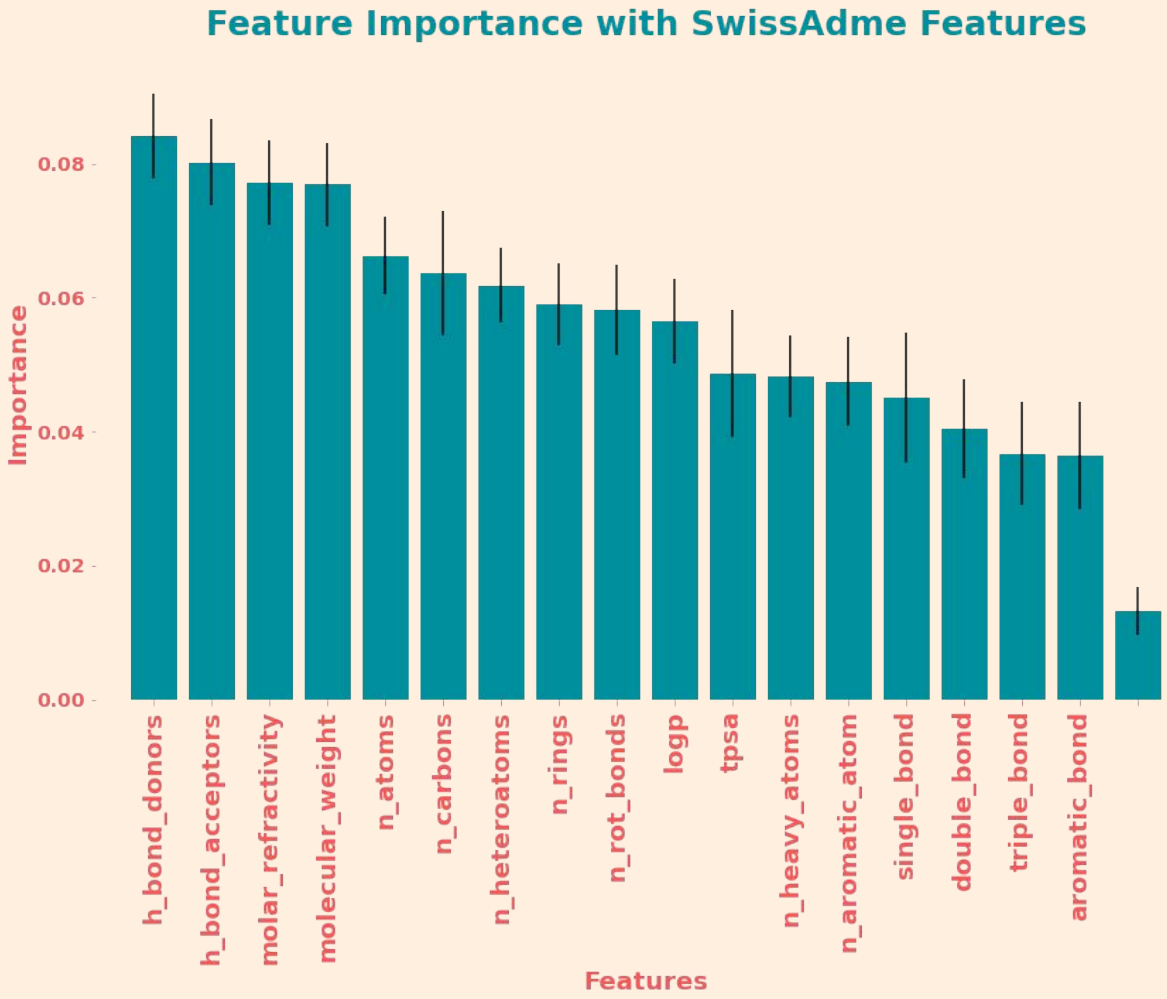
They had **50** features, and Rdkit is not the easiest to learn so I ended up with **17** of those features in the second round of well, everything.

SwissADME Feature Creation

FEATURES LIST				
			n_atoms	n_bonds
single_bond	n_heteroatoms	logp	aromatic_bond	n_heavy_atoms
double_bond	n_rings	tpsa	Molecular_ weight'	h_bond_donors
triple_bond	n_rot_bonds	N_aromatic_ atom	Molar_ refractivity	H_bond_ acceptors

EDA


Feature Importance with Fingerprint Features
Importance
0.020
0.015
0.010
0.005
0.000
12 15 31 11 46 26 27 1 53 13 14 45 24 20 35 42 32 63 40 25 54 34 48 59 37 19 8 7 6 21 57 28 62 4 3 52 50 49 18 51 41 47 23 38 60 9 56 5 58 55 29 44 30 2 43 61 16 33 0 10 22 39 17 36
Features


Feature Importance with SwissAdme Features
Importance
0.00
0.02
0.04
0.06
0.08
h_bond_donors
h_bond_acceptors
molar_refractivity
molecular_weight
n_atoms
n_carbons
n_heteroatoms
n_rings
n_rot_bonds
logp
tpsa
n_heavy_atoms
n_aromatic_atom
single_bond
double_bond
triple_bond
aromatic_bond
Features

Modeling

Markov Chain Neural Network

Hashed Morgan Fingerprint Markov Chain Neural Network

Tuned to: Accuracy Then Recall	CYP2c19	CYP2c9	CYP2d6	CYP1a2	CYP3a4
AUC	0.48	0.46	0.52	0.52	0.58
Accuracy	0.95	0.92	0.94	0.94	0.94
Dropout	0	0	0	0	0

SwissFeatures
Markov Chain Neural Network

Tuned to: Accuracy Then Recall	CYP2c19	CYP2c9	CYP2d6	CYP1a2	CYP3a4
AUC	0.42	0.60	0.56	0.53	0.49
Accuracy	0.94	0.93	0.96	0.95	0.92
Dropout	0	0	0	0	0

SUC
Gridsearched for Parameter Tuning

Hashed Morgan Fingerprint Gridsearched Tuned SUC

Weighted avg	CYP2c19	CYP2c9	CYP2d6	CYP1a2	CYP3a4
precision	0.87	0.88	0.88	0.87	0.88
recall	0.93	0.94	0.94	0.90	0.94
F1 score	0.90	0.91	0.91	0.93	0.91

SwissFeatures Gridsearched Tuned SUC

Weighted AUC	CYP2c19	CYP2c9	CYP2d6	CYP1a2	CYP3a4
precision	0.91	0.88	0.90	0.88	0.88
recall	0.91	0.93	0.92	0.94	0.90
F1 score	0.91	0.91	0.91	0.91	0.89

Wrap up

GOALS

01

Tailored Medicine

Create tailored treatment plans based on cytochrome isoenzyme activity

02

Efficiency + Efficacy

With machine learning, medicine can be quickly tailored to the patient

03

Pharmacokinetics

These models can also aide in creating new drugs that utilize these specific enzymes.

04

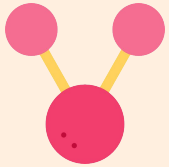
Patient Outcome

Overall patient outcome will be improved, with timeliness, and more effective treatment.

Make this into a useable app

Asked my doctor what he thought, and he thought it would be an incredibly useful tool and transformative for medicine.

What's Next?



PRODUCTIONIZE CODE

Code is not production ready. Need to convert to scripts.

Also add the rest of the SwissADME features.

TRAIN ON MORE RECENT DATA

New medications are being made every day. Keeping our project up to date will keep it relevant and useful.

HELP IAN ADD MORE TO PROJECT

There is still a lot of work to be done on the adme-pred-py project, though this is a HUGE chunk of it.

Thanks!

Resources pt.1

Modern treatment diversity issues:
https://www.sciencedaily.com/releases/2008/10/081015132108.htm
https://press.uchicago.edu/Misc/Chicago/213095.html
https://www.ncbi.nlm.nih.gov/books/NBK220337/
https://www.theguardian.com/lifeandstyle/2019/nov/13/the-female-problem-male-bias-in-medical-trials

General resources on CYP450 gene expression/affect on drug metabolism
https://www.nature.com/articles/6500462
https://www.nature.com/articles/6500285
https://www.nature.com/articles/gim2007123

SwissADME - Data gathering:
http://www.swissadme.ch
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2912882/
https://www.nature.com/articles/nbt.1581
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4477067/
https://pubmed.ncbi.nlm.nih.gov/26400175/

Resources pt.2

Data detangling:
https://depth-first.com/articles/2019/01/11/extended-connectivity-fingerprints
https://www.rdkit.org/docs/RDKit_Book.html/
https://www.researchgate.net/publication/51509126_Transfer_learning_for_cytochrome_P450_isozyme_selectivity_prediction
https://www.cheminformania.com/fetch-the-sherif-i-found-a-fingerprint/

General Research:
https://www.aafp.org/afp/2007/0801/p391.html

Toxicity
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7431904/
https://academic.oup.com/jat/article/29/7/590/731303
https://www.hindawi.com/journals/bmri/2015/971982/
https://www.sciencedirect.com/science/article/pii/S1752928X16300051

Design and images:
https://pubchem.ncbi.nlm.nih.gov/gene/CYP2C19/human(to get to rcsb for each molecule)
https://www.rcsb.org/3d-view/2HI4/1 (for 3d molecule images)
https://www.flaticon.com/
https://www.freepik.com/
https://slidesgo.com