B.TECH MINI PROJECT

Under the guidance of

Prof Amit Mankodi

OBJECTIVE

- The project aims to predict the runtime of a C language code that simulates the propagation of Plasma in a 2D space using Machine Learning techniques. The goal is to estimate the runtime on different CPUs with various configurations.
- The code has a set of parameters that can be adjusted to test the runtime. The code needs to be tested on different CPUs with a diverse set of configurations to aid in predicting the runtime.
- The initial steps involve cleaning and analyzing the runtime data. It’s crucial to identify the dominant factors that greatly affect the runtime, which can assist in designing a custom CPU optimized for this specific application.

SCOPE

- Predict the runtime of the evolution and propagation of Plasma code.
- Use Machine Learning Techniques to predict the runtime.
- Identify dominant factors which affect the runtime performance of the code.
- For performing the above tasks, analyze and clean the input data in a suitable manner to feed the model.
- Generate the dataset by testing the code on different systems each with a different set of configurations.

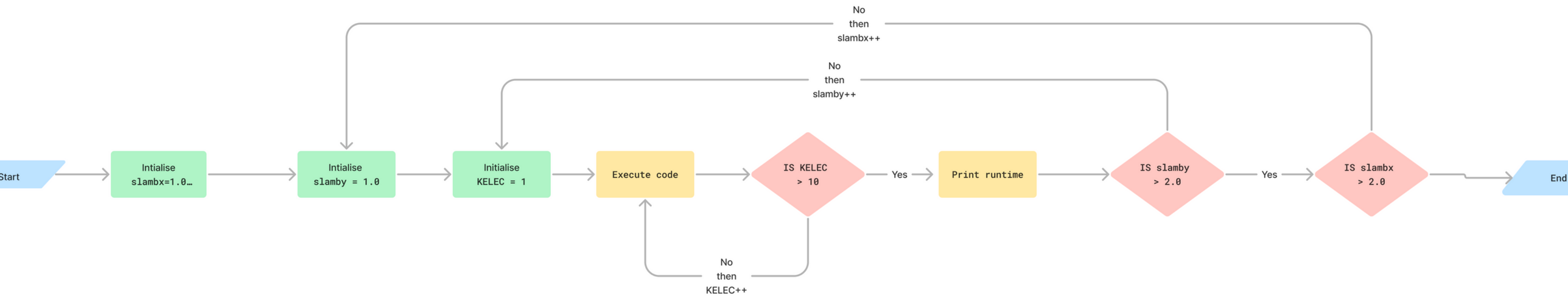
ASSUMPTIONS

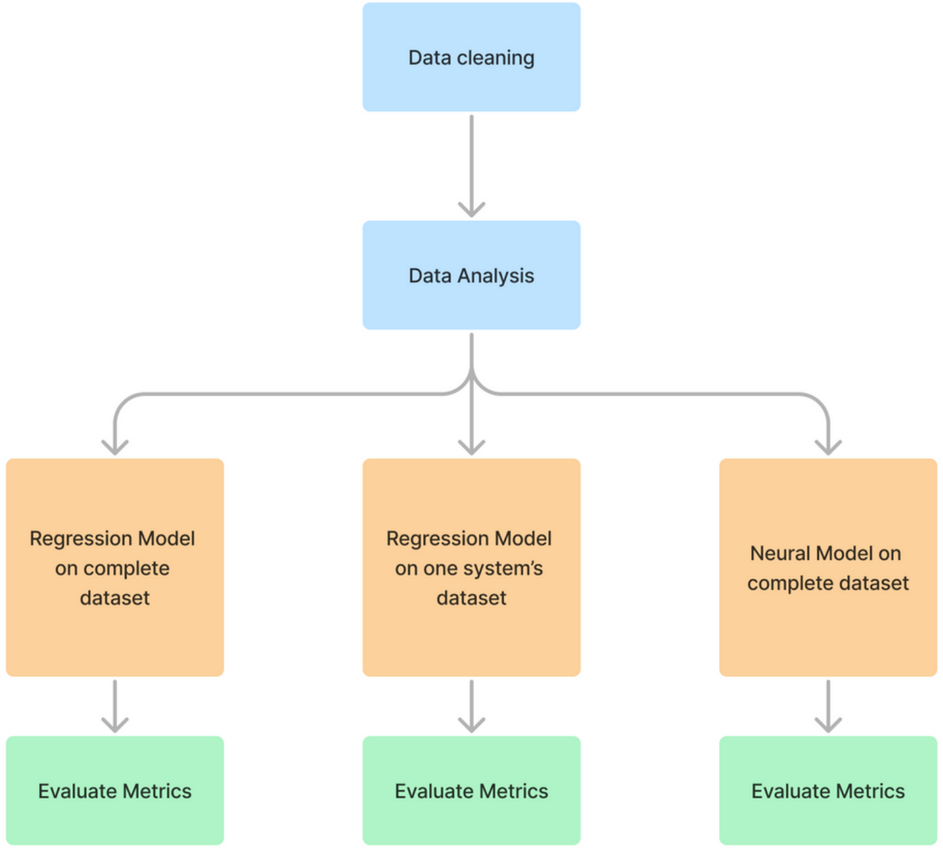
- The dataset is compiled on idle systems with almost no processes running in the background except the necessary system processes.
- It is assumed that changing the operating system would not affect the runtime or does not require any kind of optimization.
- The code uses a variable named KELEC, which is proportional to the time of propagation of Plasma in space. This variable was fixed at a constant value of 10, to speed up the process of generating the dataset. This means that for all the data points, the duration of propagation of Plasma was fixed.

TOOLS AND TECHNOLOGIES

- C Language
- Shell commands
- Python and related libraries (Numpy, Pandas, Matplotlib, Seaborn)
- Machine Learning and related libraries (Scikit-learn)
- Basics of Deep Learning with Neural Networks (Keras)
- Google Colab Notebook

WORKFLOW

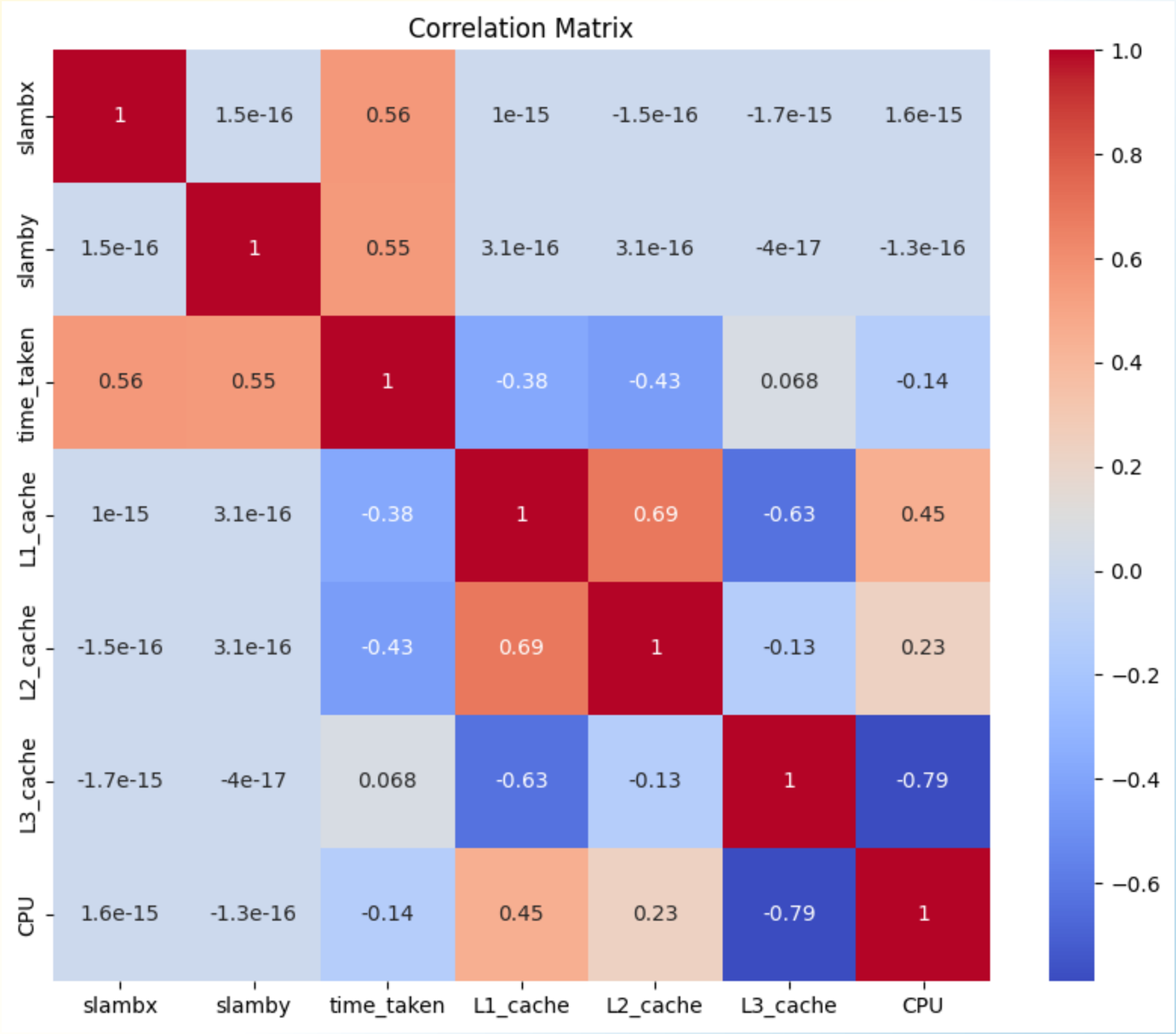

No
then
slambx++
No
then
slamby++
Start
Intialise
slambx=1.0…
Intialise
slamby = 1.0
Initialise
KELEC = 1
Execute code
IS KELEC
> 10
Yes
Print runtime
IS slamby
> 2.0
Yes
IS slambx
> 2.0
End
No
then
KELEC++


Data cleaning
Data Analysis
Regression Model on complete dataset
Regression Model on one system's dataset
Neural Model on complete dataset
Evaluate Metrics
Evaluate Metrics
Evaluate Metrics

RESULTS

	R2 score	RMSE	MAPE
Linear Regression on complete dataset	0.9138	69.7945	7.6136
Linear Regression on one system’s dataset	0.9562	32.3571	5.5828

	R2 score	RMSE	MAPE
2 layers + 200 epochs	-0.8025	300.9320	38.1465
3 layers + 200 epochs	0.2214	197.7833	26.9871
4 layers + 200 epochs	0.1777	203.2548	28.8034
2 layers + 300 epochs	-0.1150	236.6839	29.7857
3 layers + 300 epochs	0.2873	189.2215	**24.3515**
4 layers + 300 epochs	0.2197	197.9977	26.0832
2 layers + 400 epochs	-8.1909	679.5374	94.0803
3 layers + 400 epochs	**0.3303**	**183.4266**	25.2099
4 layers + 400 epochs	-8.1906	679.5247	94.0780


Correlation Matrix
slambx
slamby
time_taken
L1_cache
L2_cache
L3_cache
CPU
1
1.5e-16
0.56
1e-15
-1.5e-16
-1.7e-15
1.6e-15
1.5e-16
1
0.55
3.1e-16
3.1e-16
-4e-17
-1.3e-16
0.56
0.55
1
-0.38
-0.43
0.068
-0.14
1e-15
3.1e-16
-0.38
1
0.69
-0.63
0.45
-1.5e-16
3.1e-16
-0.43
0.69
1
-0.13
0.23
-1.7e-15
-4e-17
0.068
-0.63
-0.13
1
-0.79
1.6e-15
-1.3e-16
-0.14
0.45
0.23
-0.79
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

FUTURE SCOPE

- The parameter KELEC can also be taken into consideration as a parameter while performing the tests and generating results.
- This project uses the L1 cache, L2 cache, L3 cache, and the CPU of the system as features. But more properties like processor speed and disk I/O speed can be included.
- The serial code can be parallelized to utilize the power of parallel processing in cores and threads.
- A new version of the code which dynamically varies the area of interest based on the propagation of plasma has become available. This reduces the computation which needs to be carried out, and hence can be further analyzed.

THANK YOU

Prepared by
Hemang Joshi (202001212)