Central University of
Technology, Free State

A Prediction Model for Malaria: An Ensemble of Machine Learning and Hydrological Drought Indices.

By Muthoni Masinde

Mrs. Paulina Phoobane

Prof. Joel Botai

Presentation Schedule

- Background and Context
- Problem Statement and Objectives
- Methodology & Results
- Discussion, Conclusion and Future work

Background information and Context (1)

- Infectious diseases claim the lives of many people each year globally and Africa continues to bear the burden of morbidity.
- Most of the existing health systems that are developed to address infectious diseases and other health issues, face multiple challenges.
- Recently, innovations in the realm of big climate data and machine learning have gained popularity especially in health.
- Consequently, to prevent and control the disease outbreaks, several prediction tools have been developed.
- When it comes to malaria transmission/outbreak, we hypothesise that drought indices could play a role.

Background information and Context (2)

- Malaria accounted for over 93% of the global malaria incidence (213 million of the 228 million) reported in 2018 (Masinde, 2020).
  - Over 50% of the cases occurred in the following 6 countries: Nigeria (25%), Democratic Republic of Congo (12%), Uganda (5%), Cote d'Ivoire (4%), Mozambique (4%) and Niger (4%).
- Concerted efforts to eradicate malaria have largely succeeded e.g. WHO's Global Technical Strategy for Malaria 2016-2030
  - Climate change is negating some of the successes – cases and geographical coverage of malaria could double by 2050. (W. Cella et al.,2019)
- There are other efforts towards developing formal models for predicting malaria based on climatic conditions
  - E.g., by the Intergovernmental Panel on Climate Change's (IPCC)
- Relative to other countries, malaria is less common in South Africa

Background information and Context (3)

- Definition of drought is elusive; some types: meteorological, hydrological, agricultural and socio-economic
- Examples of drought indices: Standardized Precipitation Index (SPI) and Effective Drought Index (EDI) (Adisa et al., 2021)
- Effective Drought Index (EDI), is an effective hydrological drought index for computing the intensity of droughts. EDI also gives the Available Water Resources Index (AWRI).

Category	Range of Drought Index Values	
	SPI	EDI
Extremely Dry	≤ -2.0	≤ -2.0
Severely Dry	-1.5 to -1.99	-1.5 to -1.99
Moderately Dry	-1.0 to -1.49	-1.0 to -1.49
Normal	-0.99 to 0.99	-0.99 to 0.99
Moderately Wet	1.0 to 1.49	1.0 to 1.49
Severely Wet	1.5 to 1.99	1.5 to 1.99
Extremely Wet	≥ 2.0	≥ 2.0

Background information and Context (3)

- Definition of drought is elusive; some types: Meteoroidal, hydrological, agricultural and socio-economic
- Examples of drought indices: Standardized Precipitation Index (SPI) and Effective Drought Index (EDI) (Adisa et al., 2021)
- EDI is an effective hydrological drought index for computing the intensity of droughts. It considers daily water accumulation with a weighting function over time and provides Available Water Resources Index (AWRI).

Category	Range of Drought Index Values	
	SPI	EDI
Extremely Dry	≤−2.0	≤−2.0
Severely Dry	−1.5 to −1.99	−1.5 to −1.99
Moderately Dry	−1.0 to −1.49	−1.0 to −1.49
Normal	−0.99 to 0.99	−0.99 to 0.99
Moderately Wet	1.0 to 1.49	1.0 to 1.49
Severely Wet	1.5 to 1.99	1.5 to 1.99
Extremely Wet	≥2.0	≥2.0

Problem Statement and Objectives

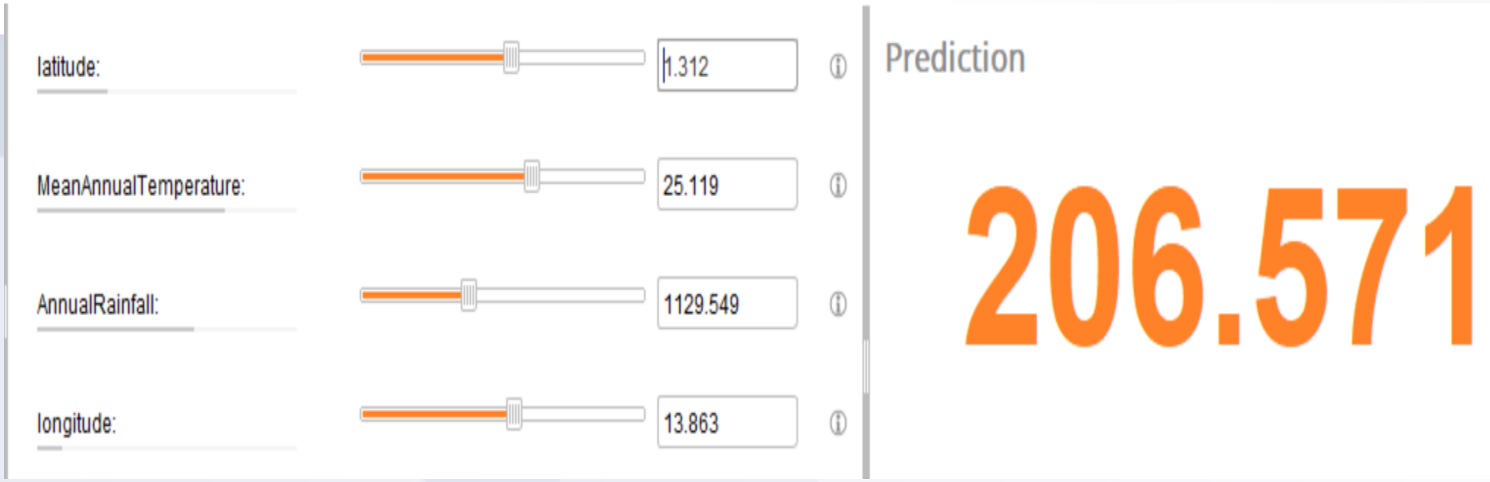
- The following research questions (in the context of Africa) are pursued in this paper:
- 1) **What is the correlation between hydrological drought indices and outbreaks of malaria?**
- 2) **What machine learning algorithms are capable of efficiently and effectively monitoring and predicting the correlation in question 1 above?**
- The main objective of this research was demarcated as: to develop a machine learning algorithm for predicting outbreaks of malaria using drought indices and historical outbreaks of malaria.
- A case study of the malaria outbreaks in Limpopo Province in South Africa.

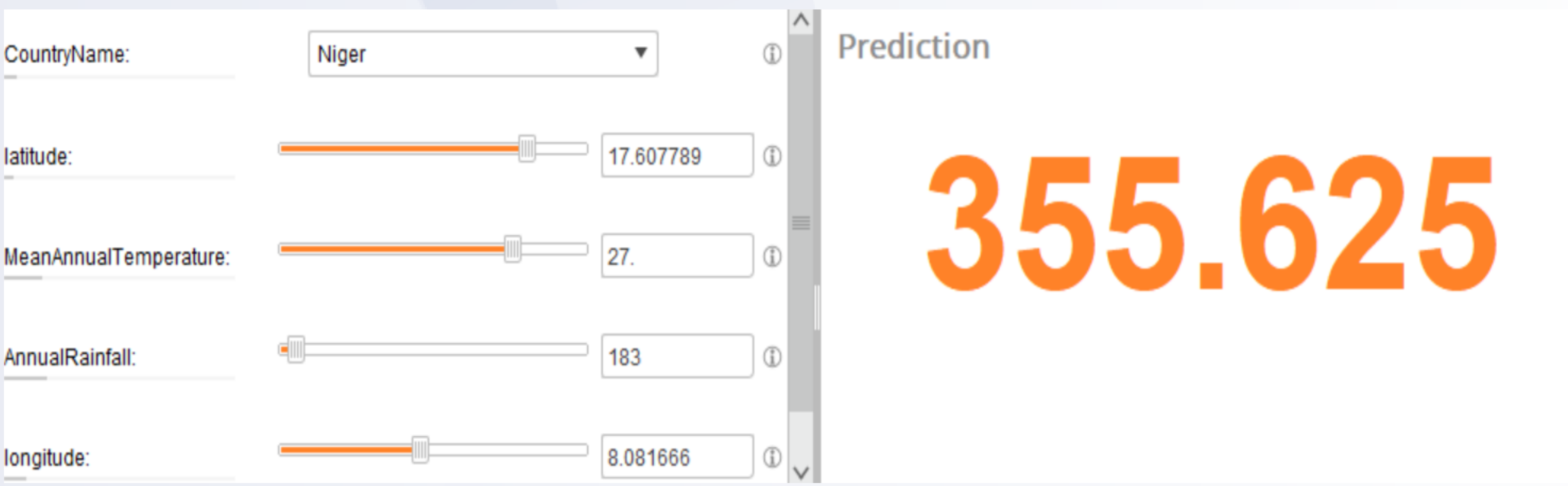
Problem Statement and Objectives

- Downscaling prior models developed with sparse data on Africa
- 1) historical incidences of malaria (measured by the number of malaria cases per 1000 population at risk per year) - http://apps.who.int/gho/data/node.gswcah
- Historical weather data consisting of rainfall and temperature - https://climateknowledgeportal.worldbank.org/download-data
- Used sing the Auto Model of RapidMiner based on six dimensions machine learning classification performance matrix
- Combining both classification and *computation* time performances put Gradient Boosted Trees, Random Forest, Decision Tree and Deep Learning as the 4 top performing algorithms respectively and Fast Large Margin as the worst performing algorithm. A new dataset was used for evaluation rusting in 0.06% to 16.43% Relative Error

Central University of Technology, Free State

Problem Statement and Objectives



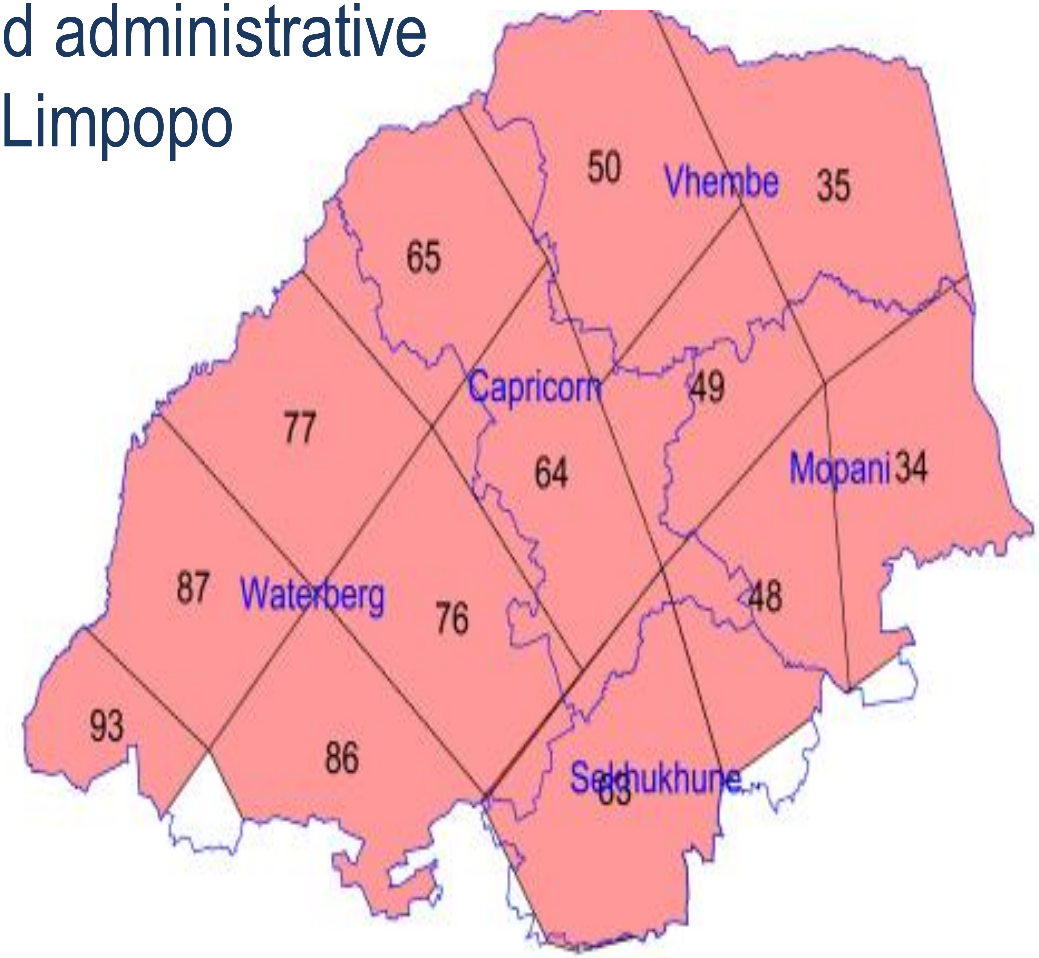


Study Area (1)

Mozambique
Botswana
Limpopo
Namibia
Gauteng
North West
Swaziland
Mpumalanga
Free State
KwaZulu-Natal
Northern Cape
Lesotho
South-Africa
www.freeworldmaps.net
Eastern Cape
Western Cape

Study Area (2)

- Climate and administrative districts in Limpopo



Datasets (1)

1. Monthly malaria incidence data per district;
   - Sourced: Limpopo province of South Africa - period:1998 to 2019.
2. Monthly weather data consisting monthly rainfall/precipitation sourced from South Africa Weather Service
   - Used to compute of EDI and AWRI

- 20 years data was used for model training and 1 year data for evaluation.

Central University of Technology, Free State

Datasets (2)

lalaria incidents (Reported Cases per Month

Capricor	Greater S	Mopani	Vhembe	Waterbei	Grand To
13	4	88	244	44	393
23	10	135	562	48	778
7	3	109	312	24	455
16	1	49	269	38	373
9	6	22	101	19	157
3		12	17		32
	1	6	9	1	17
1		5	39	9	54
3		49	141	13	206

Effective Drought Index (EDI)

Date	Dis35EDI	Dis48EDI	Dis49EDI	Dis50EDI	Dis63EDI	Dis64EDI	Dis65EDI
15 01 1998	-0.8	-0.24	0.21	-0.15	-0.01	-1.14	-0.5
15 02 1998	-1.19	-0.66	-0.57	-0.82	-0.29	-1.02	-0.94
15 03 1998	-1.33	-0.97	-0.74	-1.06	-1.21	-1.03	-0.91
15 04 1998	-1.45	-0.99	-0.79	-1.23	-1.08	-1.08	-1.09
15 05 1998	-1.49	-1.03	-0.87	-1.26	-1.16	-1.17	-1.15
15 06 1998	-1.55	-1.06	-0.94	-1.33	-1.22	-1.28	-1.18
15 07 1998	-1.56	-0.8	-0.63	-1.22	-1.17	-1.28	-1.15
15 08 1998	-1.6	-0.88	-0.71	-1.28	-1.5	-1.42	-1.23

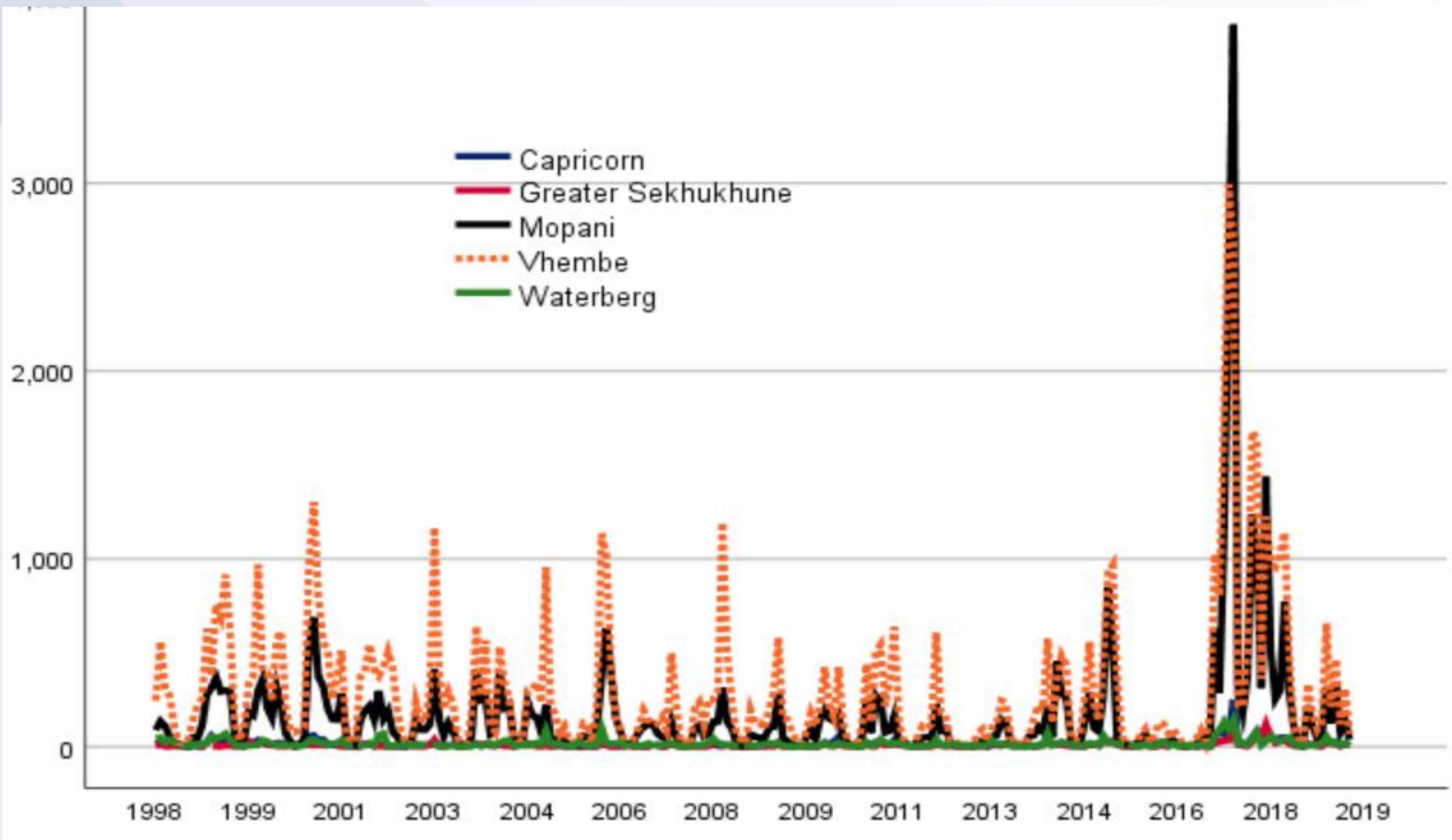
Precipitation (Rainfall)

District35	District48	District49	District50	District63	District64	District65	District76	District77	District8
89	125.7	248.8	110.5	71.1	33.5	69.9	74.2	46.7	122.
11.9	30.4	52.1	11.5	56.9	35.4	9.1	41.6	28.7	47.
13.3	24	53.2	17.9	16.3	41.4	30.8	34.2	60.1	46.
5.3	28.3	32	3.9	30.1	21	8.2	13	35	
0	0	0.2							
0	0.2	0.3							
4.5	23.9	39.8							
2.4	4.8	10.8							

Available Water Resource Index(AWRI)

Dis35AW	Dis48AW	Dis49AW	Dis50AW	Dis63AW	Dis64AW	Dis65AW	Dis76AW	Dis77AW	Dis86AV
170.2	287.1	435.1	194.6	285	130.4	148.2	218.7	144.2	292.
133.6	236.6	361	148.4	269.1	131.2	114.9	201.7	133.4	257.
114.2	201.9	324.6	129.1	219.1	139.2	117	185.2	159.4	237.
93.8	184.6	281.7	103.5	200.3	125.7	97.2	154.4	153.5	18
73.5	142.5	218.8	80.9	154.6	96.2	75.6	119.6	115.9	145.
57.9	112	172.3	63.7	121.1	75.6	59.5	94.2	91	114.
49.4	110.8	173.7	54.8	93.6	59.1	46.3	73.4	71.5	89.
39.2	86.5	139.5	44.2	70.9	45.3	35.1	56.3	55.4	68.

Malaria incidences in Limpopo



Data Processing

- Data pre-processing and feature reduction processes were done
- The variables (EDI, AWRI and precipitation) were ranked based on their correlation with malaria incidences.
- The variables with the strong correlation with malaria incidence were mostly the AWRI values.

Building the Model

- RapidMiner was used to simulate data and build the predictor models.
- Variables with higher correlations with malaria incidences in Vhembe and at provincial level were used to build the models.

3.4 Machine Learning Algorithms Performance

Central University of Technology, Free State

Table 1. Machine Learning algorithms classification Performance (Provincial data)

Model	Accuracy	AUC	Classification Error	F-Measure	Precision	Recall	Sensitivity
Decision Trees	98.7%	0%	1.3%	99.3%	98.7%	100%	100%
Deep Learning	98.7%	0%	1.3%	99.3%	98.7%	100%	100%
Fast Large Margin	98.7%	0%	1.3%	99.3%	98.7%	100%	100%
Generalized Linear Model	98.7%	0%	1.3%	99.3%	98.7%	100%	100%
Gradient Boosted Trees	98.7%	0%	1.3%	99.3%	98.7%	100%	100%
Logistic Regression	98.7%	0%	1.3%	99.3%	98.7%	100%	100%
Naïve Bayes	98.7%	0%	1.3%	99.3%	98.7%	100%	100%
Random Forest	98.7%	0%	1.3%	99.3%	98.7%	100%	100%

Machine Learning Algorithms Performance cont.

Central University of Technology, Free State

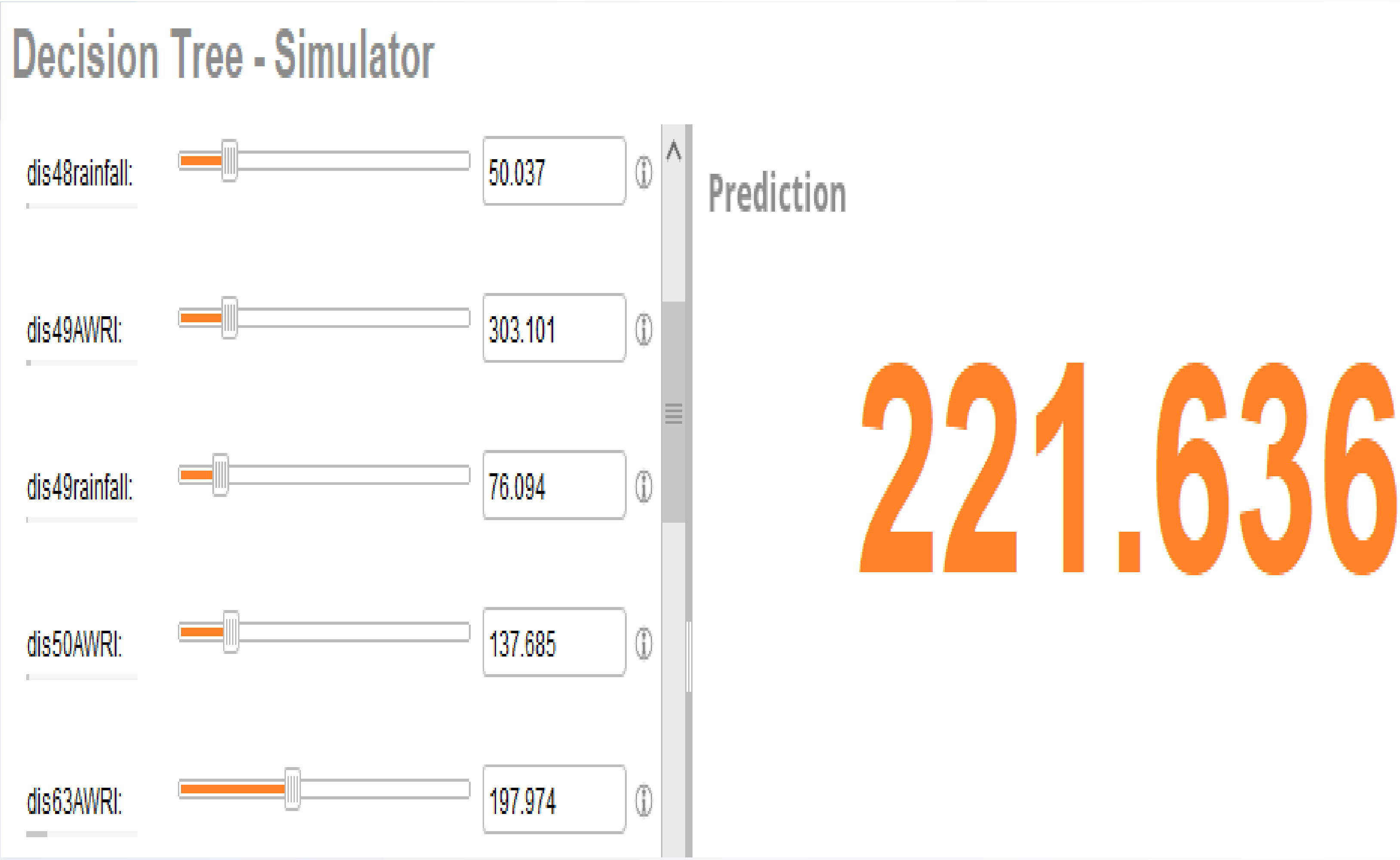
Table 2. Machine Learning algorithms classification Performance (Vhembe data)							
Model	Accuracy	AUC	Classification Error	F-Measure	Precision	Recall	Sensitivity
Decision Trees	100%	0%	0%	100%	100%	100%	100%
Deep Learning	100%	0%	0%	100%	100%	100%	100%
Fast Large Margin	error	error	error	error	error	error	error
Generalized Linear Model	100%	0%	0%	100%	100%	100%	100%
Gradient Boosted Trees	100%	0%	0%	100%	100%	100%	100%
Logistic Regression	100%	0%	0%	100%	100%	100%	100%
Naïve Bayes	100%	0%	0%	100%	100%	100%	100%
Random Forest	100%	0%	0%	100%	100%	100%	100%

Machine Learning Algorithms Performance cont.

Central University of Technology, Free State

Table 3. Machine Learning algorithms computation time Performance ranking							
Model	Total Time	Training Time (1 000 rows)	Scoring Time (1 000 rows)	Ranking-Total Time	Ranking-Training Time	Ranking-scoring time	Overall Ranking
Naive Bayes	32731,0	3135,7	951,5	5	3	6	(4) 14
Generalized Linear Model	20311,0	10182,2	757,3	3	8	4	(5)15
Logistic Regression	15004,0	7639,5	417,5	1	6	1	(2)8
Deep Learning	22747,0	9984,5	679,6	4	7	3	(4)14
Decision Tree	17840,0	2934,1	466,0	2	2	1	(1)6
Random Forest	41618,0	651,2	825,2	6	1	5	(3)13
Gradient Boosted Trees	383287,0	3581,4	2019,4	8	4	8	(7)20
Support Vector	70159,0	4391,5	1349,5	7	5	7	(6)19

Malaria Prediction System Development



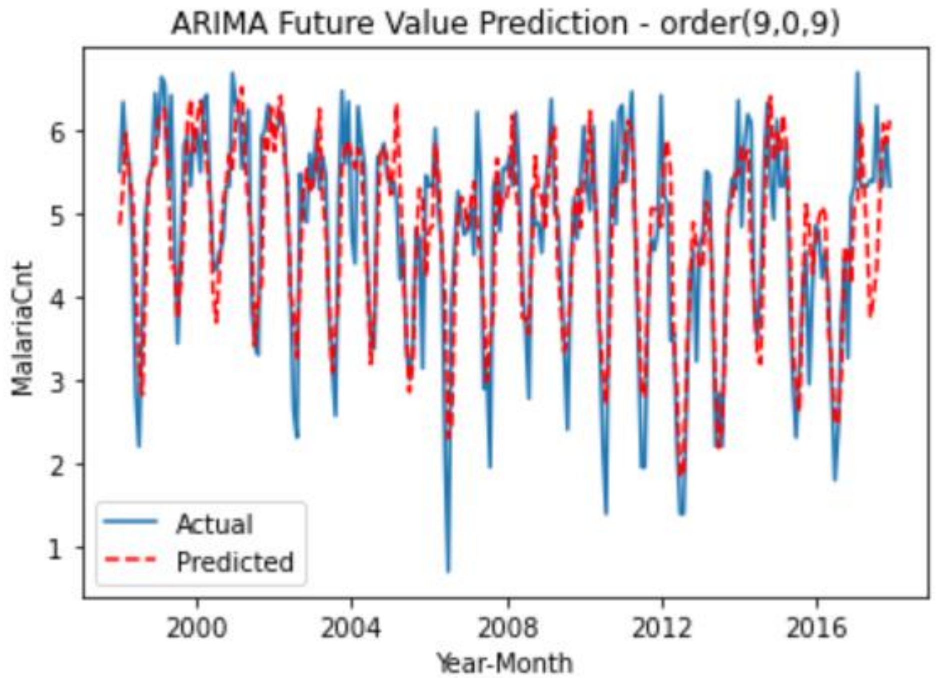
Evaluation, Discussion and Conclusion

- Even though Decision Tree, had a good performance during the training process, it did not do as good as expected during the testing.
- Research **objective 1;** it is evident that there is a very strong correlation between hydrological drought indices and malaria incidence reported in Limpopo.
- Research **objective 2;** Decision Tree Malaria Predictor System was built using in RapidMiner.
- AWRI is an excellent measure of conduciveness of malaria causing vector and it is, therefore, a good predictor for malaria incidence.
- The prediction capacity of the malaria prediction system built in this study is an excellent decision tool for stakeholders.

Challenges and Further Work

- Given the small dataset, the resulting models tend to over fit – implementation of the models in real-life was not possible.
- Further to incorporate other variables such as temperature, humidity and vegetation indices has resulted in huge errors.
- Acquisition of matching (with location of malaria incidents) remains a challenge.
- Work underway to integrate Indigenous knowledge on weather and malaria outbreak (https://www.itiki.co.za/).
- Further evaluation of the model and ascertaining the model's scalability to the other parts of South Africa, Southern Africa and Africa at large.

Future work –ARIMA model




```
AIC:  580.0379700621556
BIC:  649.6507485289955
Mean Absolute Error:  0.5787644740088534
Root Mean Squared Error:  0.7426863647955668
Durbin-Watson statistic : 1.9175769620255847
```

Future work –ARIMA model

Central University of Technology, Free State

```
In [328]: predictionsdf
Out[328]:
```

	Actual	Predicted
0	5.497168	4.857146
1	6.331502	5.303281
2	5.743003	5.951086
3	5.594711	5.401519
4	4.615121	5.215290
...	...	...
235	6.289716	4.344396
236	5.321052	5.277724
237	5.321052	6.068869
238	5.888878	5.841518
239	5.321052	6.113463

240 rows × 2 columns

```
mean_abs_percent_error = sum_errors / sum_actuals
print("Model MAPE:")

print(str(round(mean_abs_percent_error * 100)) + '%')
print()
print("Model Accuracy:")

print(str(round((1 - mean_abs_percent_error)* 100)) + '%')
```

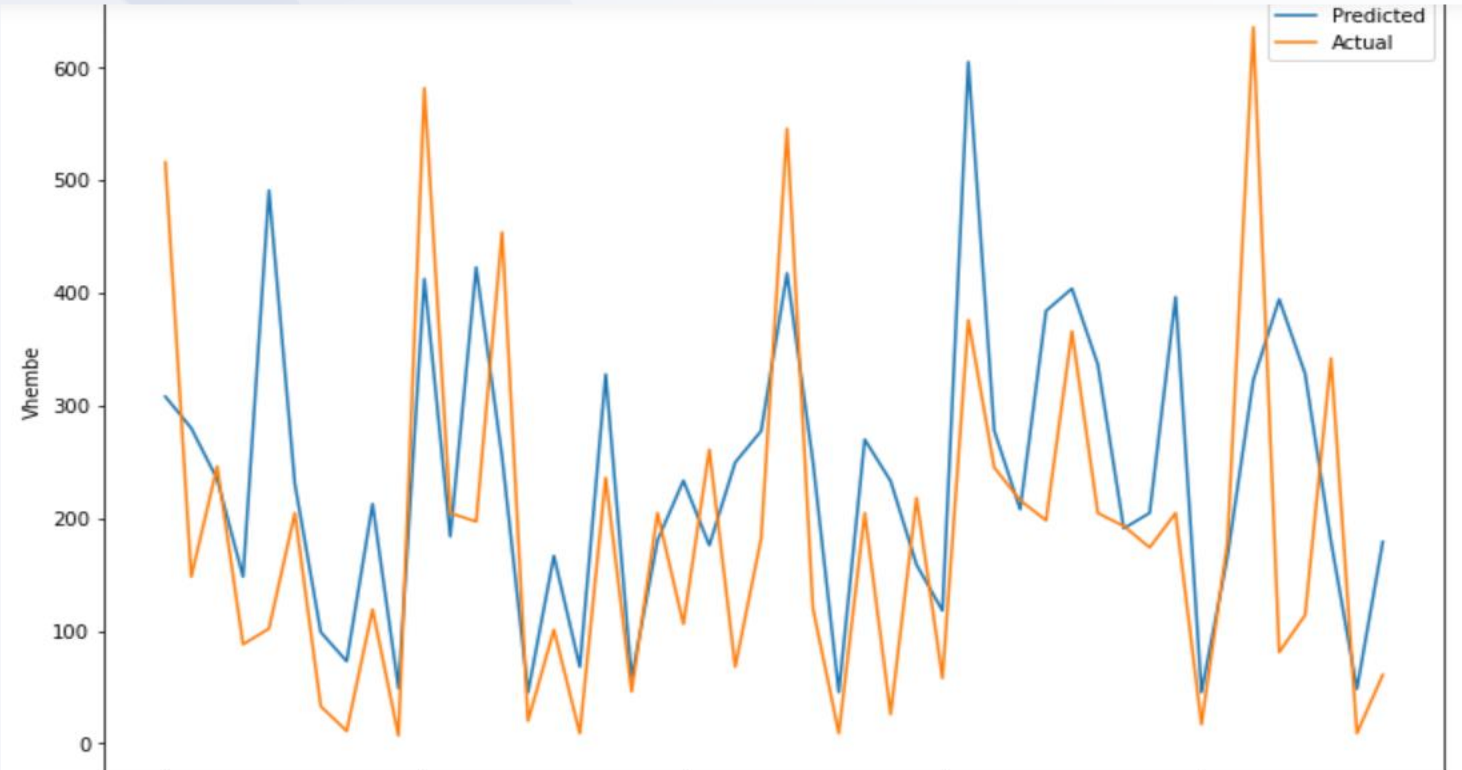
```
Model MAPE:
12%

Model Accuracy:
88%
```

Central University of Technology, Free State

Future work –ARIMA model

```
Training_score :  0.7410357286315676
Test_score :  0.17904218769388236
r_score :  0.1790421876938827
mean_squared_error :  19947.737619105745
Number of features Selected :  10
Feature Selection Technique :  F Regresion Method
```



Future work – ICMHEWS

- Incorporate the Malaria Predictor in the Integrated Climate-driven Multi-Hazard Early Warning System (ICMHEWS)

Government of Flanders (Financier)

South African Weather Service (Host)

University of Kwa-Zulu Natal (UKZN) - Partner

Central University of Technology – Free State (CUT) – Partner

Some References

Adisa, O.M., Masinde, M. and Botai, J.O., 2021. Assessment of the dissimilarities of EDI and SPI measures for drought determination in South Africa. Water, 13(1), p.82.

Byun, H.;Wilhite, D.A. Objective quantification of drought severity and duration. J. Clim. 1999, 12, 2747–2756.

Masinde, M. (2020, March). Africa's Malaria Epidemic Predictor: Application of Machine Learning on Malaria Incidence and Climate Data. In Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis (pp. 29-37).

Phoobane, P., Masinde, M. and Botai, J., 2022. Prediction Model for Malaria: An Ensemble of Machine Learning and Hydrological Drought Indices. In Proceedings of Sixth International Congress on Information and Communication Technology (pp. 569-584). Springer, Singapore.

W. Cella et al. 2019, ‘Do climate changes alter the distribution and transmission of malaria? Evidence assessment and recommendations for future studies’, Rev. Soc. Bras. Med. Trop., vol. 52, 2019

Q & A

- Thank you

www.cut.ac.za

Central University of Technology, Free State

Prof Muthoni Masinde
Professor and Head of Department
Department of Information Technology
Faculty of Engineering, Built Environment and Information Technology

E: emasinde@cut.ac.za or muthonimasinde@gmail.com

T: +27 51 507 3091

Vision 2030

The Central University of Technology. Shaping the future through innovation as a leading African University of Technology.