Segurança e Privacidade

Trabalho prático 1

Tiago Conceição – uc2021167993

1. Homormophic Encryption

Para implementar este método de encriptação, foi utilizado a biblioteca Pyfhel que é uma das bibliotecas sugeridas, sendo selecionado o seguinte dataset selecionado weather_hourly_darksky.csv.

Neste dataset numa primeira abordagem foram selecionados os dados que correspondiam ao mês '2012-11-01' e '2013-11-01', mas devido ao número reduzido de dados que foram selecionados a análise não foi tão precisa, pois contava apenas com 1440 registos.

Dada a análise obtida, teve que ser repensada a estratégia de seleção de dados, que passou a ser todos os registos acima da data '2012-01-01', que resultou num total de 19701 registos. E nesses registos optou-se por calcular a soma das temperaturas e as médias das mesmas.

	visibility	temperature	time	pressure	apparentTemperature	windSpeed	precipType	icon	humidity	summary
1464	4.09	10.19	2012-11-25 00:00:00	992.53	10.19	5.62	rain	partly-cloudy-night	0.99	Mostly Cloudy
1465	7.45	11.67	2012-11-25 01:00:00	990.38	11.67	6.70	rain	partly-cloudy-night	0.98	Mostly Cloudy
1466	8.95	12.54	2012-11-25 02:00:00	989.61	12.54	7.97	rain	wind	0.92	Breezy and Mostly Cloudy
1467	13.50	11.39	2012-11-25 03:00:00	992.15	11.39	11.02	rain	wind	0.86	Breezy and Partly Cloudy
1468	13.50	10.26	2012-11-25 04:00:00	994.81	10.26	11.31	rain	wind	0.84	Windy and Mostly Cloudy
...	...	...	...	...	...	...	...	...	...	...
21160	12.68	7.39	2014-02-15 19:00:00	997.07	3.91	6.08	rain	partly-cloudy-night	0.74	Partly Cloudy
21161	13.78	6.56	2014-02-15 20:00:00	998.15	3.03	5.61	rain	clear-night	0.77	Clear
21162	14.31	6.47	2014-02-15 21:00:00	999.28	3.06	5.25	rain	clear-night	0.77	Clear
21163	14.31	5.96	2014-02-15 22:00:00	1000.33	2.68	4.69	rain	clear-night	0.80	Clear
21164	14.31	5.38	2014-02-15 23:00:00	1001.25	1.77	5.09	rain	clear-night	0.82	Clear

Figura 0-1 - Dataset com os dados selecionados a partir de 2012

E apesar de ser um número relativamente reduzido, o processo de encriptação ainda foi um bocado demorado pois foi selecionado a coluna da temperatura toda para ser encriptada numa tentativa de melhoria de performance para com os futuros cálculos.

Como se pode ver na imagem seguinte a imagem essa coluna está toda encriptada.

	visibility	temperature	time	pressure	apparentTemperature	windSpeed	precipType	icon	humidity	summary
1464	4.09	<Pyfhel Ciphertext at 0x7f18bc6cf840, encoding...	2012-11-25 00:00:00	992.53	10.19	5.62	rain	partly-cloudy-night	0.99	Mostly Cloudy
1465	7.45	<Pyfhel Ciphertext at 0x7f18bc4778c0, encoding...	2012-11-25 01:00:00	990.38	11.67	6.70	rain	partly-cloudy-night	0.98	Mostly Cloudy
1466	8.95	<Pyfhel Ciphertext at 0x7f18bc6cb400, encoding...	2012-11-25 02:00:00	989.61	12.54	7.97	rain	wind	0.92	Breezy and Mostly Cloudy
1467	13.50	<Pyfhel Ciphertext at 0x7f18bc6cb4c0, encoding...	2012-11-25 03:00:00	992.15	11.39	11.02	rain	wind	0.86	Breezy and Partly Cloudy
1468	13.50	<Pyfhel Ciphertext at 0x7f18bc6cbc40, encoding...	2012-11-25 04:00:00	994.81	10.26	11.31	rain	wind	0.84	Windy and Mostly Cloudy
...	...	...	...	...	...	...	...	...	...	...
21160	12.68	<Pyfhel Ciphertext at 0x7f189bb3a400, encoding...	2014-02-15 19:00:00	997.07	3.91	6.08	rain	partly-cloudy-night	0.74	Partly Cloudy
21161	13.78	<Pyfhel Ciphertext at 0x7f189bb3a440, encoding...	2014-02-15 20:00:00	998.15	3.03	5.61	rain	clear-night	0.77	Clear
21162	14.31	<Pyfhel Ciphertext at 0x7f189bb3a480, encoding...	2014-02-15 21:00:00	999.28	3.06	5.25	rain	clear-night	0.77	Clear
21163	14.31	<Pyfhel Ciphertext at 0x7f189bb3a4c0, encoding...	2014-02-15 22:00:00	1000.33	2.68	4.69	rain	clear-night	0.80	Clear
21164	14.31	<Pyfhel Ciphertext at 0x7f189bb3a500, encoding...	2014-02-15 23:00:00	1001.25	1.77	5.09	rain	clear-night	0.82	Clear

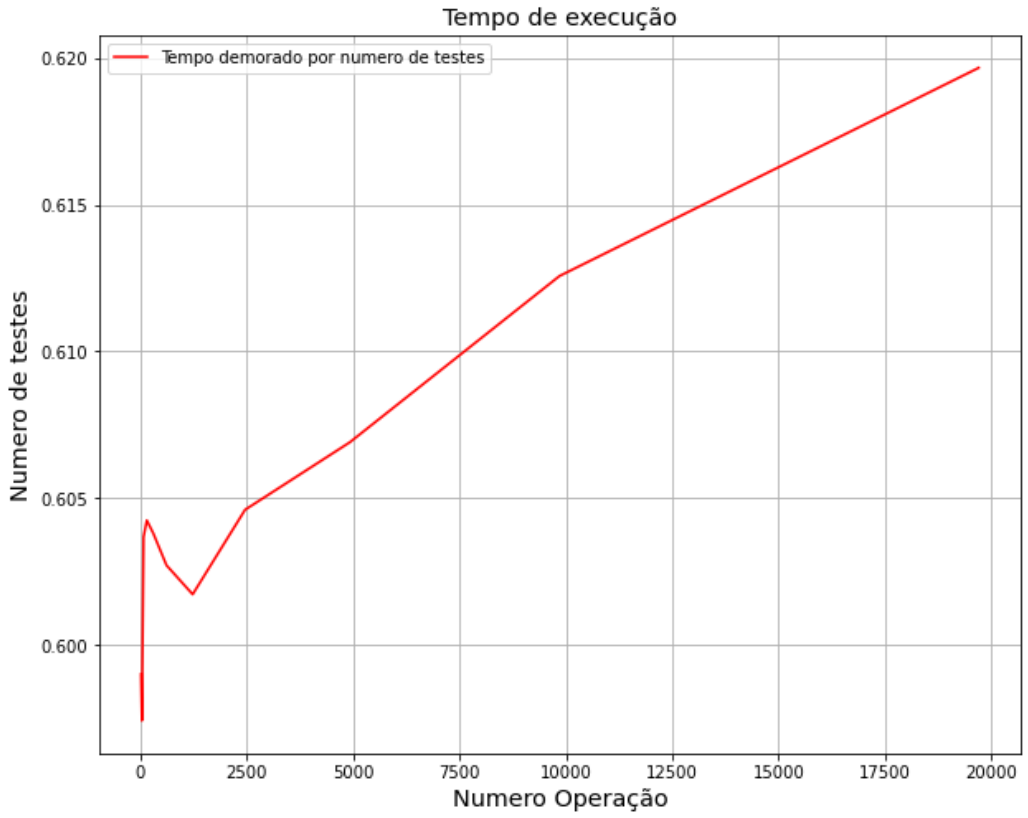
Figura 0-2 - HE aplicado à linha da temperatura

De seguida passou-se à contabilização do tempo demorado por cada registo, começando por 17 registos a ser analisados até 19701 registos, e nota-se que o tempo vai progressivamente aumentando como se pode ver na figura seguinte.

	0	1	2	3	4	5	6	7	8	9	10
n_testes	19701.00000	9850.00000	4925.00000	2462.00000	1231.00000	616.00000	308.0000	154.00000	77.00000	38.00000	19.00000
numero_op	0.61966	0.61257	0.60691	0.60461	0.60173	0.60272	0.6038	0.60425	0.60368	0.59743	0.59901

Figura 0-3 - Dados de tempo demorado por número de testes

Como se pode analisar o tempo está abaixo de 1 segundo, mas rapidamente passará essa marca se tratássemos de mais registos.



Este tipo de encriptação tem algumas desvantagens como o elevado custo de computação exigido e como alguns erros de calculo que são observados posteriormente nos ficheiros de cálculos o que pode trazer consequências a quem queira fazer uma análise futura ao dataset.