Testing ADABoost and Bagging

Para esta tarea se utilizaron las características extra extraídas en la entrega previa:

- Cantidad de menciones que contiene el tweet (numérica).
- Es retweet o no (nominal).
- Cantidad de ligas que contiene el tweet (numérica).
- Cantidad de palabras que contiene el tweet (numérica).
- 4 atributos de personalidad provistos por la api de indico.

Bayesian Networks

Se realizaron pruebas utilizando diferentes algoritmos de búsqueda:

- K2, que implementa hill climbing con restringido con el orden de valores..
- Hill Climbing.
- Repeated Hill Climbing.
- Simulated Annealing.
- Tabu search.
- TAB.

También se utilizaron variaciones en los parámetros de cada algoritmo, en algunos, por ejemplo, se varió en número de nodos padres. Los esultados obtenidos son los siguientes:

BNSETAB-18F.csv — **0.69722**
3 hours ago by Carlos A
Bayesian Network Simple estimator, search algorithm TAB: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNSE3Par-18F.csv — **0.71282**
3 hours ago by Carlos A
Bayesian Network Simple estimator, search algorithm Tabu Search 3 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNSESimAnn-18F.csv — **0.72982**
3 hours ago by Carlos A
Bayesian Network Simple estimator, search algorithm Simulation Annealing: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNSE3Par-18F.csv 0.70014

3 hours ago by Carlos A

Bayesian Network Simple estimator, search algorithm Repeated Hill Climbing 3max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNHC7ParMetaCost-18F.csv 0.74996

3 hours ago by Carlos A

MetaCost [0, 1, 1.75, 0] with Bayesian Network Simple estimator, search algorithm Hill Climbing 7 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNHC7Par-18F.csv 0.72315

3 hours ago by Carlos A

Bayesian Network Simple estimator, search algorithm Hill Climbing 7 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNHC3Par-18F.csv 0.71309

3 hours ago by Carlos A

Bayesian Network Simple estimator, search algorithm Hill Climbing 3 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNHC1Par-18F.csv 0.68998

3 hours ago by Carlos A

Bayesian Network Simple estimator, search algorithm Hill Climbing 1 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNMBASE1ParBayST-18F.csv 0.68998

3 days ago by Carlos A

Bayesian Network MBA estimator, search algorithm k2 2 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNK2SE7ParBayST-18F.csv 0.70276

3 days ago by Carlos A

Bayesian Network simple estimator search algorithm k2 7 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNK2SE2ParBaySTwMB-18F.csv 0.71262
3 days ago by Carlos A
Bayesian Network simple estimator search algorithm k2 2 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNK2SE2ParBayST-18F.csv 0.71262
3 days ago by Carlos A
Bayesian Network simple estimator search algorithm k2 2 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

BNK2SE1ParBayST-18F.csv 0.69871
3 days ago by Carlos A
Bayesian Network simple estimator search algorithm k2 1 max parent: 5 sentiment 5 emotion 4 personality, number of mentions, retweet and count of links and word

Conclusiones

Algunos de los resultados fueron similares a los obtenidos con Random Forest, por este motivo experimentaré combinando la red bayesiana con método de resampling, bagging y cost sensitive.