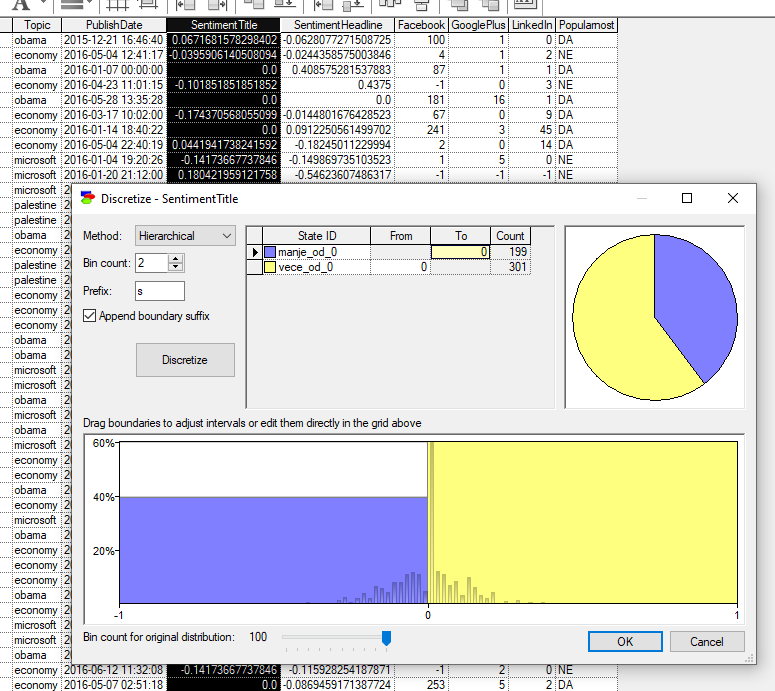
# TREĆI DIO PROJEKTNOG ZADATKA – BAYESOVE MREŽE

## Uvoz podataka u GeNiE

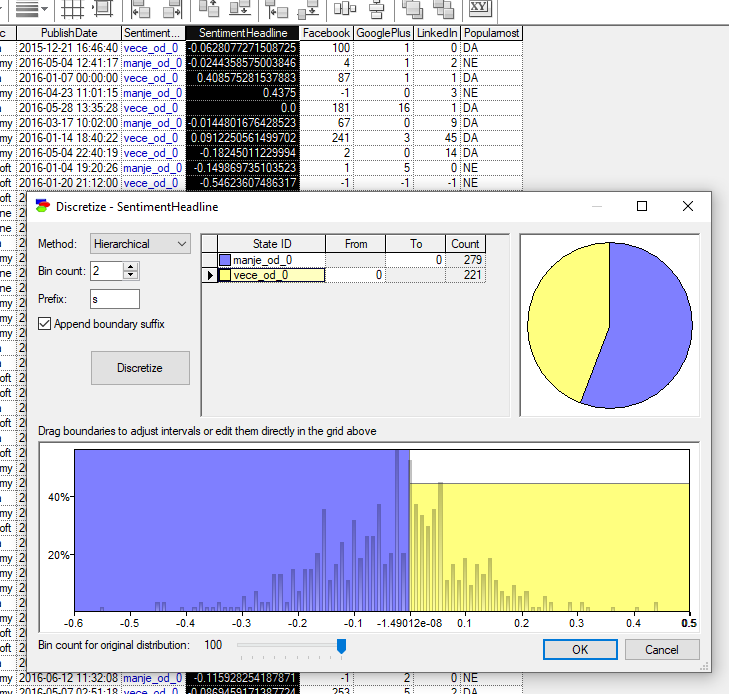
Podaci u program GeNiE uveženi su iz datoteke „Novosti\_Bayes.csv“ koja se razlikuje od data seta korištenog u 1. i 2. projektom zadatku za atribut, odnosno ciljnu varijablu „Popularnost“ te još dva irelevantna atributa koji su izbačeni (redni broj i ID).

## Diskretizacija numeričkih varijabli

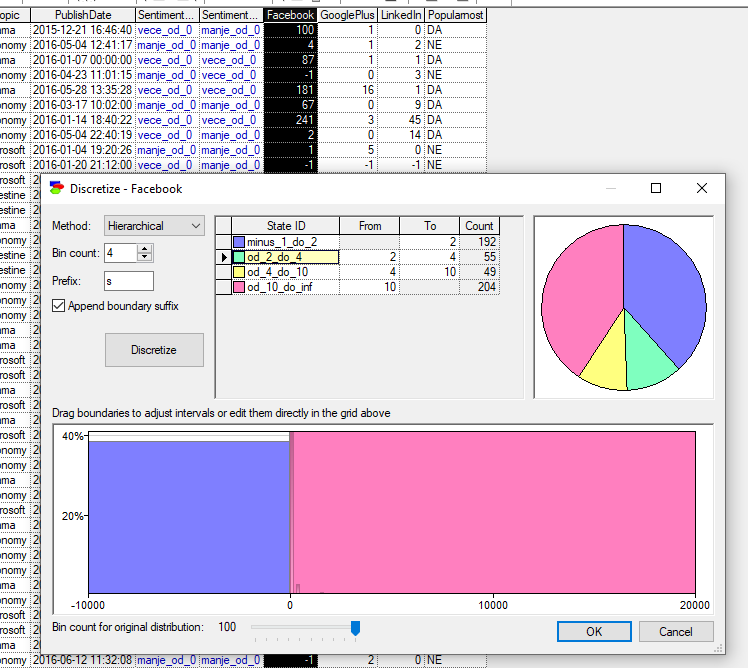
„SentimentTitle“:



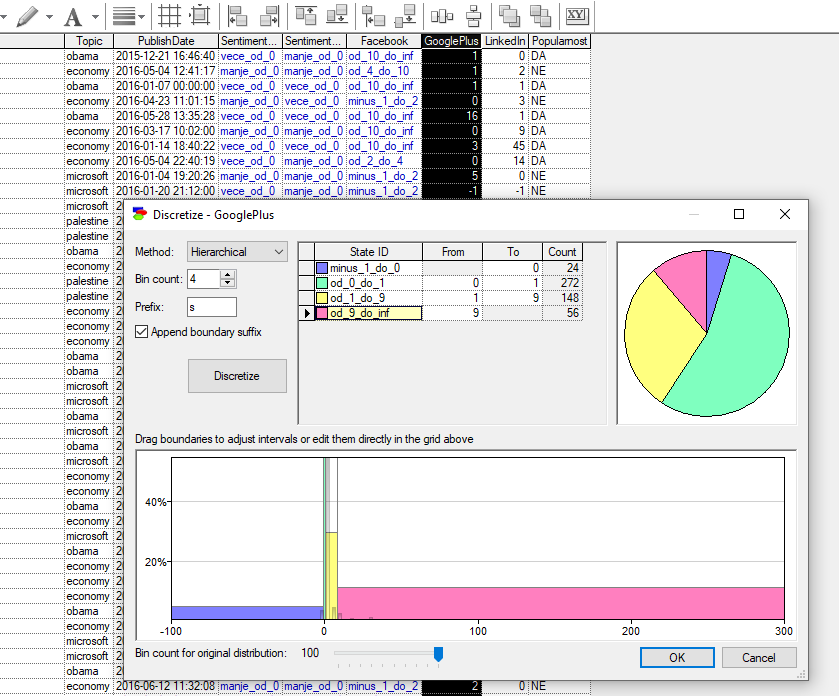
„SentimentHeadline“:



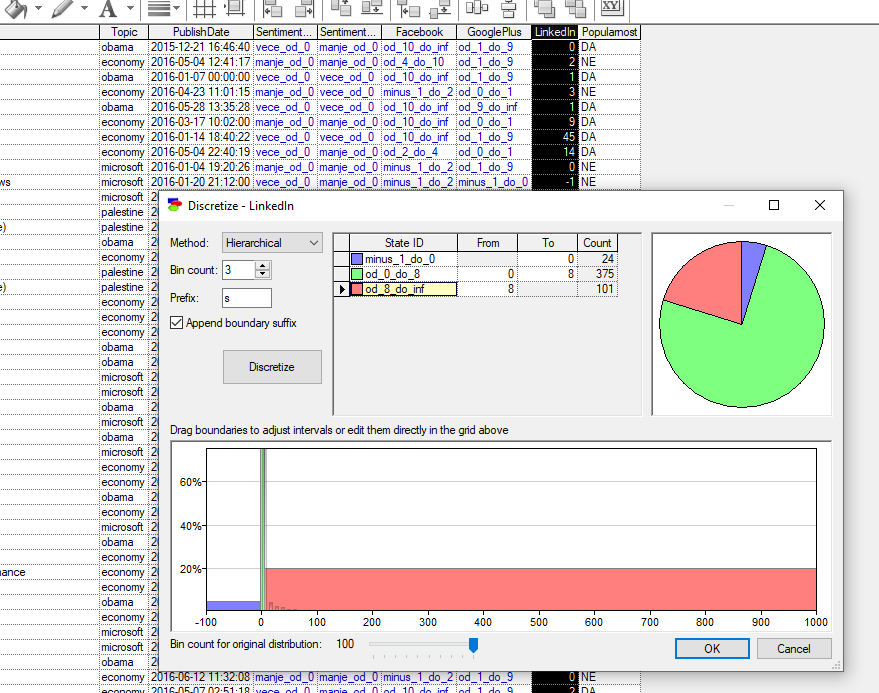
„Facebook“:



„GooglePlus“:

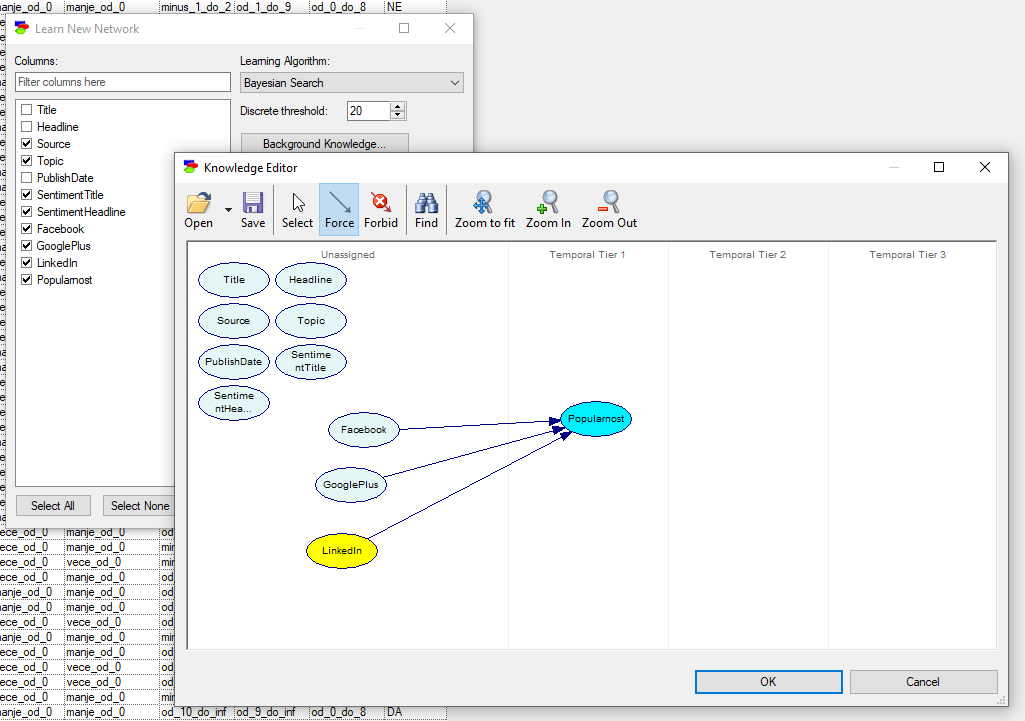


„LinkedIn“:



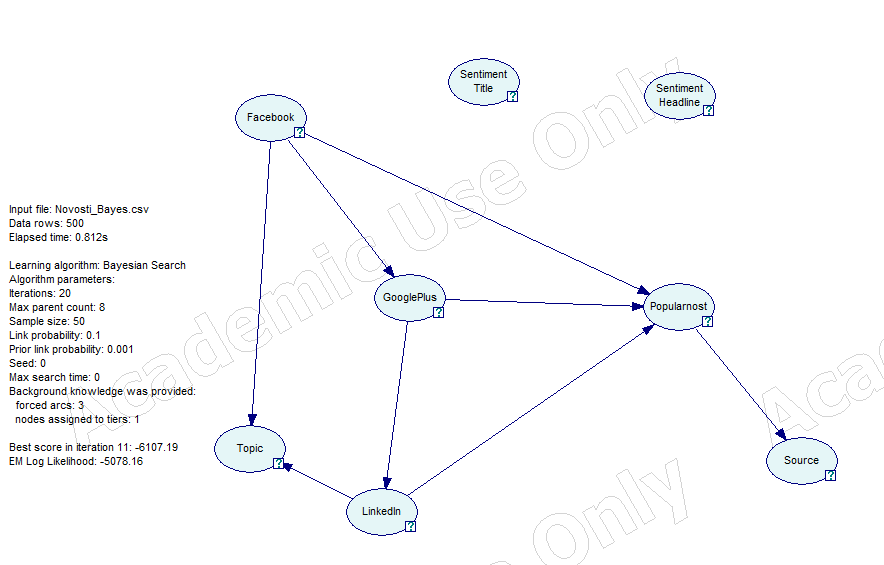
## Kreiranje i treniranje modela

Varijable „Facebook“, „GooglePlus“ i LinekdIn“ imaju najvišu informacijsku vrijednost te će biti povezane s ciljnom varijablom:

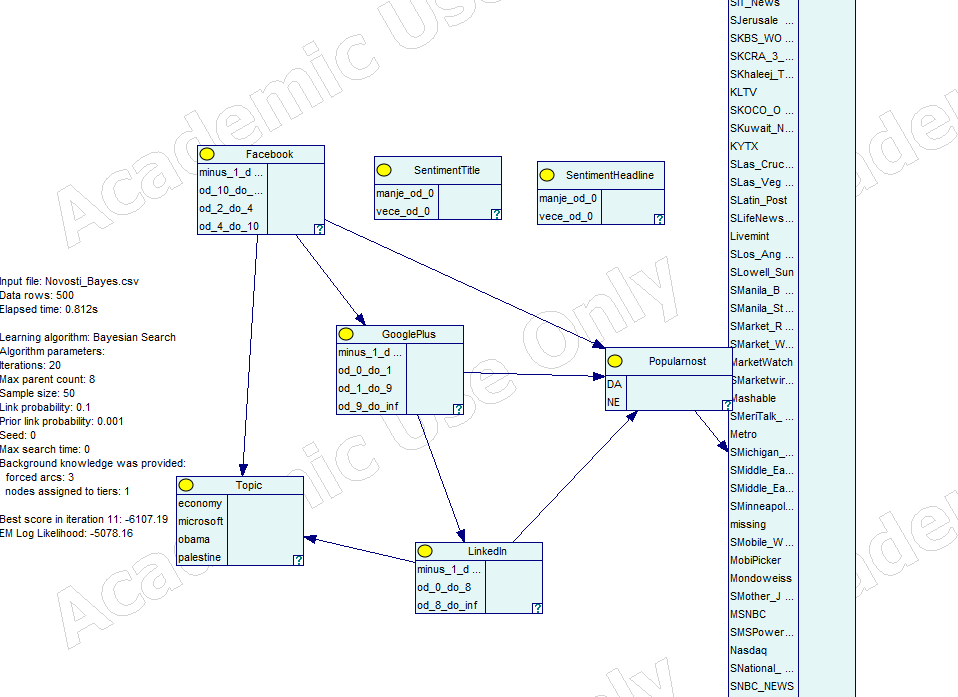


Atributi „Headline“, „Title“ i „PublishDate“ su izostavljeni jer se program iz nekog razloga ruši ako su navedeni atributi uključeni. A i pretpostavka je da „Headline“ i „Title“ ne utječu na ciljnu varijablu sa svojom vrijednošću (semantikom itekako, ali u kontekstu izračuna te varijable su samo unikatni niz znakova).

Nakon treniranja modela stvoreno je 5 novih veza:



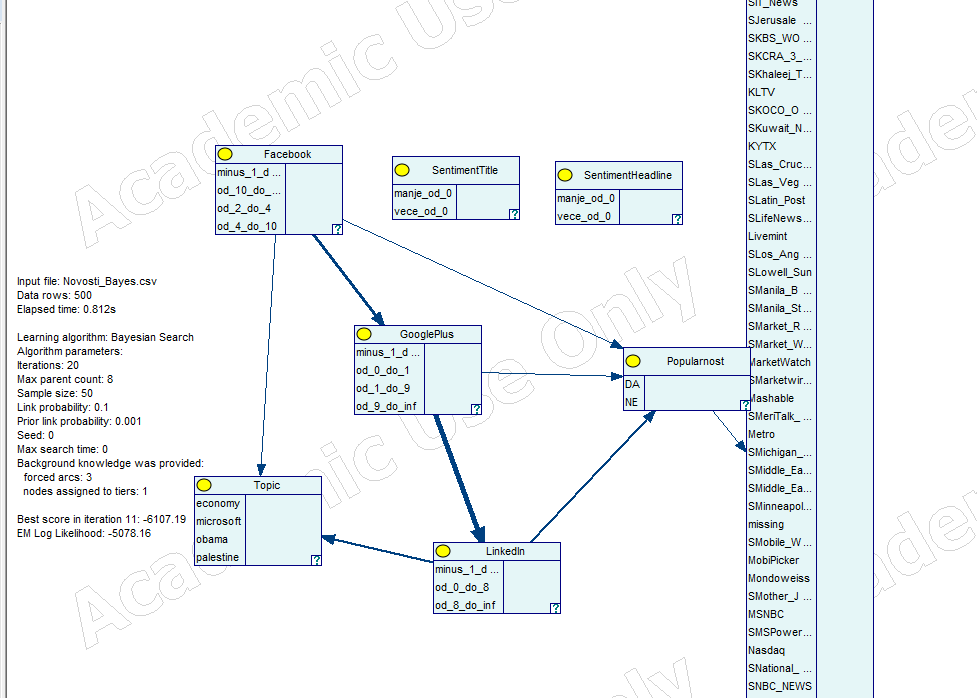
Stvorene su veze između „Facebook“ i „GooglePlus“, „Facebook“ i „Topic“, „GooglePlus“ i „LinkedIn“, „LinkedIn“ i „Topic“ te između „Popularnost“ i „Source“



Iz navedenih rezultata, pretpostavka je da „Facebook“ ima utjecaj na popularnost vijesti na „GooglePlus“, a „GooglePlust“ na popularnost vijesti na „LinkedIn“. Također je li vijest popularna ili ne, ima utjecaja o kojem se izvoru novosti radi. Atributi „Facebook“ i „LinkedIn“ imaju utjecaja i na kategoriju vijesti „Topic“ (economy, microsoft, obama, palestine).

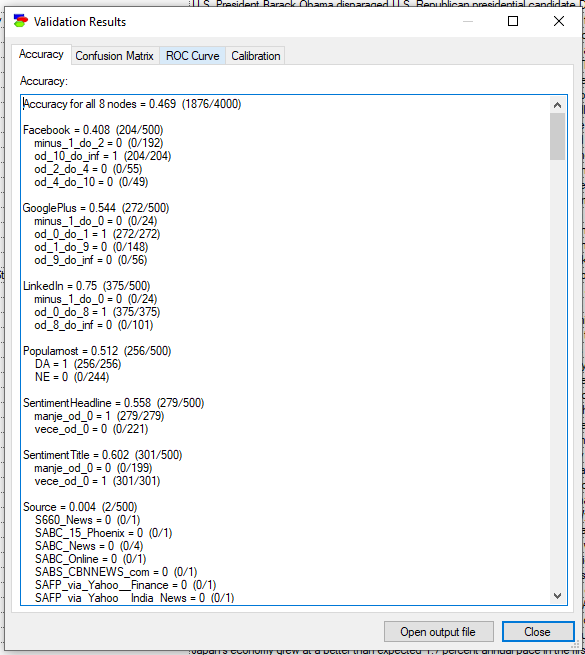
## Prikaz jačina veza/utjecaja

Vidljivo je da postoji jači utjecaj „Facebook“ na „GooglePlus“ te jači utjecaj „LinkedIn“ na „Topic“ i „Popularnost“. Najjači utjecaj je „GooglePlus“ na „LinkedIn“.

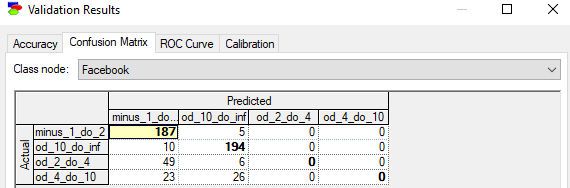


## 

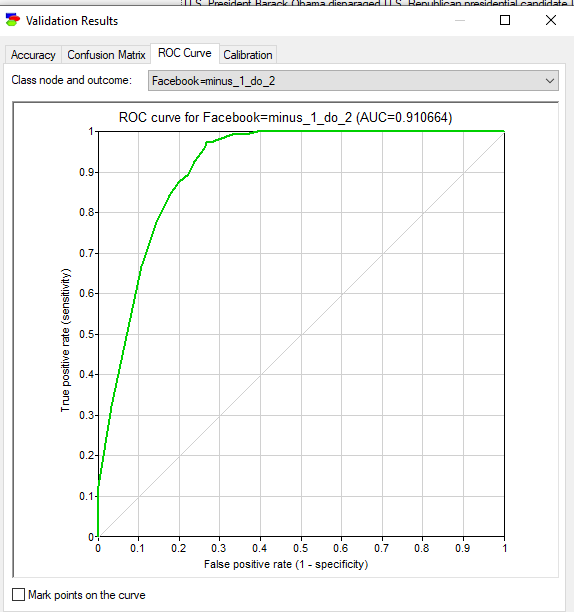
Iz rezultata skoriranja vidljivo je da je ukupna točnost predikcije za svih 8 čvorova nešto manja od 50% dok je najveća točnost predikcije za čvor „LinkedIn“ i iznosi 75%.



Matrica zbunjenosti prikazuje broj točno i pogrešno klasificiranih zapisa. Primjer za čvor Facebook:

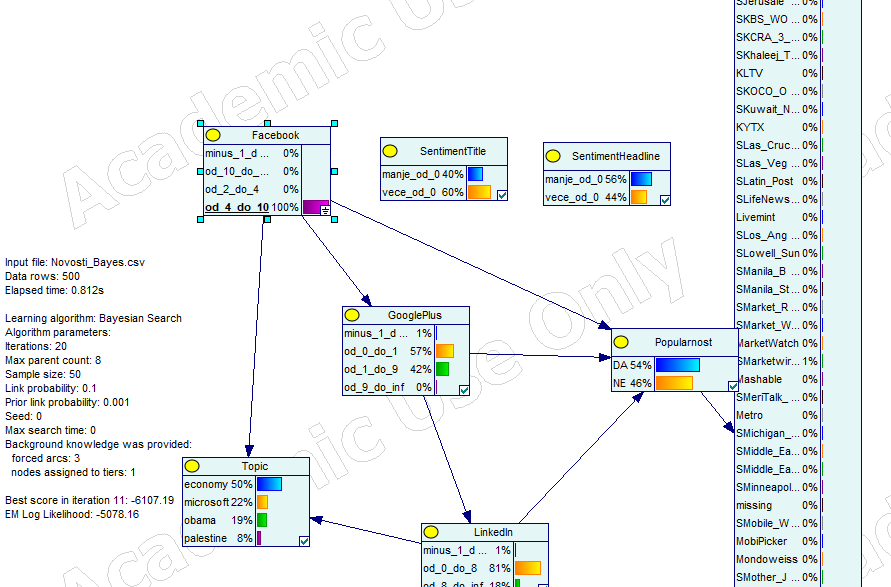


ROC krivulja za „Facebook“ kategorija (-1,2):

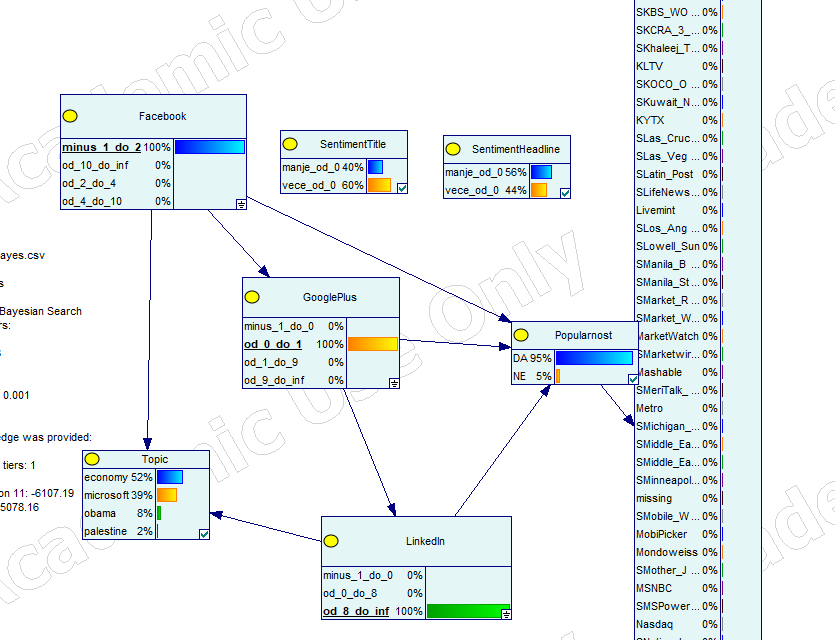


## Nekoliko scenarija

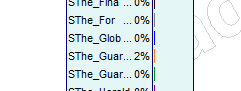
1. „Evidence“ (vjerojatnost = 1) postavljen na atribut „Facebook“ kategorija „od\_4\_do\_10“ pokazuje da je vjerojatnost da vijest bude popularna 54%:



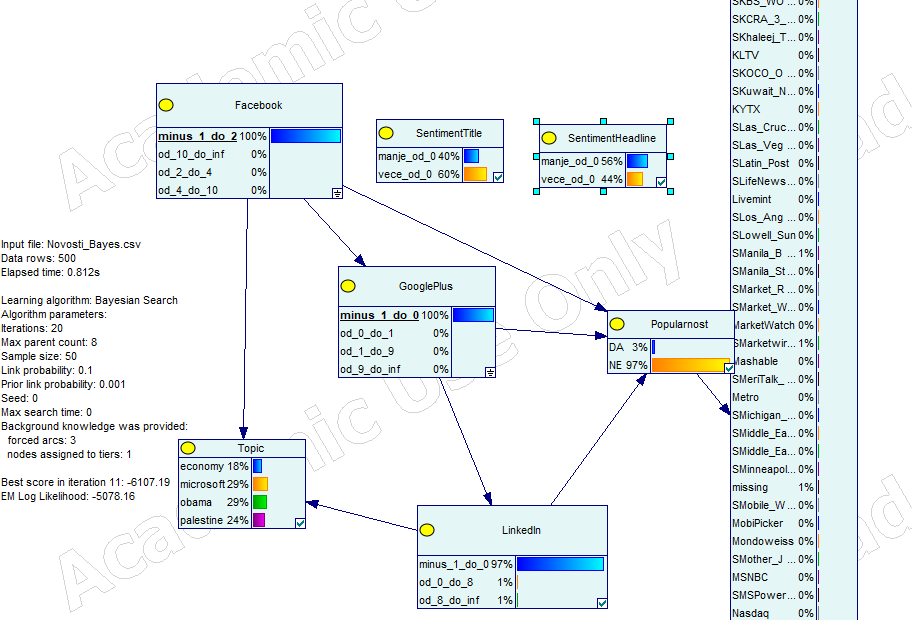
1. „Evidence“ postavljen na „LinkedIn“ - „od\_8\_do\_inf“, „Facebook“ - „minus\_1\_do\_2“ te „GooglePlus“ – „od\_0\_do\_1“ pokazuje da je tada vjerojatnost 95% da će vijest biti popularna te da će biti iz područja ekonomije. Dakle vijest iako nepopularna na Facebooku i GooglePlus će biti popularna i iz područja ekonomije. Također se može primijetiti različita raspodjela sentimenta za Title i Headline što može govoriti da ljudi nakon što pročitaju Title(manje informacija) promjene sentiment nakon što pročitaju Headline(više informacija).



Pri čemu je najveća šansa da izbor bude „Guardian“:



1. „Evidence“ postavljen na „Facebook“ - „od\_1\_do\_2“, „GooglePlus“ - „minus\_1\_do\_0“ pokazuje vjerojatnost od 97% da će vijest biti popularna na „LinkedIn“ te vjerojatnost od 97% da će biti nepopularna. Također je vjerojatnost 29% da će biti tematika biti „microsoft“ ili „obama“:



1. „Evidence“ postavljen na „Topic“ – „palestine“ pokazuje vjerojatnost od 77% da će vijest biti nepopularna:

