

YOUTUBE SNIMCI

Seminarski rad iz istraživanja podataka 1
Matematički fakultet

Studenti: Nevena Soldat 504/2017
Đorđe Stanojević 224/2015

Uvod

U okviru ovog rada bavićemo se analizom skupa podataka o Jutjub snimcima i komentarima prikupljenim u Velikoj Britaniji i Sjedinjenim Američkim Državama. Prvo ćemo opisati podatke, obaviti pretprocesiranje, a zatim različitim algoritmima pronaći pravila pridruživanja i komentarisati dobijene rezultate. Sva obrada podataka izvršena je u IBM SPSS Modeleru.

Podaci – analiza i pretprocesiranje

Podaci koje ćemo koristiti su preuzeti sa adrese <https://www.kaggle.com/datasnaek/youtube>. Imamo četiri skupa podataka – USvideos i GBvideos koji sadrže podatke o video snimcima u SAD/ Velikoj Britaniji, za kojima slede UScomments i GBcomments koji sadrže reakcije i komentare na te video snimke. Prva dva skupa imaju sledeće attribute:

video_id	Identifikator snimka
Title	Naziv snimka
channel_title	Naziv kanala na kojem se snimak nalazi
category_id	Kategorija
Tags	Oznake
Views	Broj pregleda
Likes	Broj odobravanja
Dislikes	Broj neodobravanja
comment_total	Broj komentara
thumbnail_link	Link ka slici snimka
Date	Datum postavljanja

U sledeća dva skupa imamo:

video_id	Identifikator snimka
comment_text	Komentari
Likes	Broj odobravanja komentara
Dislikes	Broj neodobravanja komentara

Pored ovih podataka, imamo dva dodatna json fajla – GB_category_id i US_category_id, u kojima se nalazi opis kategorija video snimaka:






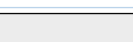




Etag	Elektronska oznaka
Items	Stavke
Kind	Vrsta

Prilikom analize podataka, primetili smo da skupovi Videos i Comments u svom izvornom obliku imaju nekoliko loše formatiranih redova koji su uticali na prikazivanje ostalih slogova. Naime, na nekim slogovima separatori su bili izostavljeni, nekima su nedostajali navodnici, neki su sadržali podatke iz pogrešnih tabela ili su imali višak/manjak polja. Kako je ovo sprečavalo ispravno učitavanje datoteka u značajnoj meri, odlučili smo da ih ručno ispravimo. Nakon učitavanja, korišćenjem čvora Data Audit dobili smo razne informacije o podacima i saznali nešto više o nedostajućim vrednostima. Rezultati toga se mogu videti na sledećim slikama i to redom prvo za GBComments pa za USComments.

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
video_id		Categorical	--	--	--	--	--	--	384290
comment_...		Categorical	--	--	--	--	--	--	384274
likes		Continuous	0	60630	2.615	159.270	208.397	--	384290
replies		Continuous	0	521	0.203	5.639	73.006	--	384289

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
video_id	Categorical	--	--	Never	Never	Fixed	100	384290	0	0	0	0
comment_text	Categorical	--	--	Never	Never	Fixed	99.996	384274	0	16	16	0
likes	Continuous	48	129 None	Never	Never	Fixed	100	384290	0	0	0	0
replies	Continuous	145	226 None	Blank & Null Values	Blank & Null Values	Fixed	100	384289	1	0	0	0

Slika 1: Informacije o skupu podataka GBcomments

Audit	Quality	Annotations							
Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
 video_id		 Categorical	--	--	--	--	--	--	450539
 comment_...		 Categorical	--	--	--	--	--	--	450524
 likes		 Continuous	0	48772	2.815	160.389	145.083	--	450539
 replies		 Continuous	0	529	0.240	6.513	62.292	--	450539

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
video_id	Categorical	--	--	--	Never	Fixed	100	450539	0	0	0	0
comment_text	Categorical	--	--	--	Never	Fixed	99.997	450524	0	14	15	0
likes	Continuous	42	161	None	Never	Fixed	100	450539	0	0	0	0
replies	Continuous	169	333	None	Never	Fixed	100	450539	0	0	0	0

Slika 2: Informacije o skupu podataka UScomments

Vidimo da se nedostajuće vrednosti pojavljuju samo u jednom slogu u skupu GBComments, tako da bismo ovaj jedan slog mogli odbaciti, ipak korišćenjem čvora za izračunavanje nedostajućih vrednosti je ovaj problem uklonjen zamenom nedostajuće vrednosti prosečnom.

Prilikom analiziranja .json fajlova, videli smo da u jednom imamo kategoriju više nego u drugom. U pitanju je kategorija “Nonprofits & Activism” (category_id=29), u okviru US_category_id fajla, što se može uočiti na fotografiji u okviru desne tabele:

Table (1 Fields, 31 records) #1	Table (1 Fields, 32 records)																																																																																																																																		
<table> <tr><th>Field</th><th>Annotations</th></tr> <tr><td>1</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Film & Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>2</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Autos & Vehicles', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>3</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Music', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...</td></tr> <tr><td>4</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Pets & Animals', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>5</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sports', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...</td></tr> <tr><td>6</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Short Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...</td></tr> <tr><td>7</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Travel & Events', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>8</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Gaming', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>9</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Videoblogging', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>10</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'People & Blogs', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>11</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>12</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Entertainment', 'kind': 'youtube#videoCategory', 'etag': 'Ym...</td></tr> <tr><td>13</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'News & Politics', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>14</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Howto & Style', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...</td></tr> <tr><td>15</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Education', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>16</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Science & Technology', 'kind': 'youtube#videoCategory', 'eta...</td></tr> <tr><td>17</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>18</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>19</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Action/Adventure', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>20</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Classics', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>21</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>22</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Documentary', 'kind': 'youtube#videoCategory', 'etag': 'Ym...</td></tr> <tr><td>23</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Drama', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>24</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Family', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>25</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Foreign', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>26</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Horror', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>27</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sci-Fi/Fantasy', 'kind': 'youtube#videoCategory', 'etag': 'Ym...</td></tr> <tr><td>28</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Thriller', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>29</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shorts', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>30</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shows', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>31</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Trailers', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> </table>	Field	Annotations	1	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Film & Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...	2	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Autos & Vehicles', 'kind': 'youtube#videoCategory', 'etag': 'Y...	3	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Music', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...	4	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Pets & Animals', 'kind': 'youtube#videoCategory', 'etag': 'Y...	5	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sports', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...	6	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Short Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...	7	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Travel & Events', 'kind': 'youtube#videoCategory', 'etag': 'Y...	8	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Gaming', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	9	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Videoblogging', 'kind': 'youtube#videoCategory', 'etag': 'Y...	10	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'People & Blogs', 'kind': 'youtube#videoCategory', 'etag': 'Y...	11	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	12	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Entertainment', 'kind': 'youtube#videoCategory', 'etag': 'Ym...	13	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'News & Politics', 'kind': 'youtube#videoCategory', 'etag': 'Y...	14	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Howto & Style', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...	15	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Education', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	16	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Science & Technology', 'kind': 'youtube#videoCategory', 'eta...	17	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	18	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...	19	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Action/Adventure', 'kind': 'youtube#videoCategory', 'etag': 'Y...	20	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Classics', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	21	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	22	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Documentary', 'kind': 'youtube#videoCategory', 'etag': 'Ym...	23	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Drama', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	24	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Family', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	25	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Foreign', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	26	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Horror', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	27	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sci-Fi/Fantasy', 'kind': 'youtube#videoCategory', 'etag': 'Ym...	28	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Thriller', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	29	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shorts', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	30	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shows', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	31	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Trailers', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	<table> <tr><th>Field</th><th>Annotations</th></tr> <tr><td>1</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Film & Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>2</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Autos & Vehicles', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>3</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Music', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...</td></tr> <tr><td>4</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Pets & Animals', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>5</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sports', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...</td></tr> <tr><td>6</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Short Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...</td></tr> <tr><td>7</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Travel & Events', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>8</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Gaming', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>9</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Videoblogging', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>10</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'People & Blogs', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>11</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>12</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Entertainment', 'kind': 'youtube#videoCategory', 'etag': 'Ym...</td></tr> <tr><td>13</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'News & Politics', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>14</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Howto & Style', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...</td></tr> <tr><td>15</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Education', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>16</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Science & Technology', 'kind': 'youtube#videoCategory', 'eta...</td></tr> <tr><td>17</td><td>['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Nonprofits & Activism', 'kind': 'youtube#videoCategory', 'eta...</td></tr> <tr><td>18</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>19</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>20</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Classics', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>21</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Action/Adventure', 'kind': 'youtube#videoCategory', 'etag': 'Y...</td></tr> <tr><td>22</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>23</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Documentary', 'kind': 'youtube#videoCategory', 'etag': 'Ym...</td></tr> <tr><td>24</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Drama', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>25</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Family', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>26</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Foreign', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>27</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Horror', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>28</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sci-Fi/Fantasy', 'kind': 'youtube#videoCategory', 'etag': 'Ym...</td></tr> <tr><td>29</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Thriller', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>30</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shorts', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>31</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shows', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> <tr><td>32</td><td>['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Trailers', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...</td></tr> </table>	Field	Annotations	1	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Film & Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...	2	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Autos & Vehicles', 'kind': 'youtube#videoCategory', 'etag': 'Y...	3	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Music', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...	4	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Pets & Animals', 'kind': 'youtube#videoCategory', 'etag': 'Y...	5	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sports', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...	6	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Short Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...	7	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Travel & Events', 'kind': 'youtube#videoCategory', 'etag': 'Y...	8	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Gaming', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	9	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Videoblogging', 'kind': 'youtube#videoCategory', 'etag': 'Y...	10	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'People & Blogs', 'kind': 'youtube#videoCategory', 'etag': 'Y...	11	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	12	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Entertainment', 'kind': 'youtube#videoCategory', 'etag': 'Ym...	13	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'News & Politics', 'kind': 'youtube#videoCategory', 'etag': 'Y...	14	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Howto & Style', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...	15	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Education', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	16	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Science & Technology', 'kind': 'youtube#videoCategory', 'eta...	17	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Nonprofits & Activism', 'kind': 'youtube#videoCategory', 'eta...	18	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	19	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...	20	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Classics', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	21	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Action/Adventure', 'kind': 'youtube#videoCategory', 'etag': 'Y...	22	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	23	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Documentary', 'kind': 'youtube#videoCategory', 'etag': 'Ym...	24	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Drama', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	25	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Family', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	26	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Foreign', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	27	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Horror', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	28	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sci-Fi/Fantasy', 'kind': 'youtube#videoCategory', 'etag': 'Ym...	29	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Thriller', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	30	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shorts', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	31	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shows', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...	32	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Trailers', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...
Field	Annotations																																																																																																																																		
1	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Film & Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
2	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Autos & Vehicles', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
3	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Music', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...																																																																																																																																		
4	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Pets & Animals', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
5	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sports', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...																																																																																																																																		
6	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Short Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...																																																																																																																																		
7	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Travel & Events', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
8	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Gaming', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
9	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Videoblogging', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
10	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'People & Blogs', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
11	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
12	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Entertainment', 'kind': 'youtube#videoCategory', 'etag': 'Ym...																																																																																																																																		
13	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'News & Politics', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
14	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Howto & Style', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...																																																																																																																																		
15	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Education', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
16	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Science & Technology', 'kind': 'youtube#videoCategory', 'eta...																																																																																																																																		
17	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
18	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
19	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Action/Adventure', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
20	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Classics', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
21	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
22	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Documentary', 'kind': 'youtube#videoCategory', 'etag': 'Ym...																																																																																																																																		
23	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Drama', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
24	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Family', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
25	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Foreign', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
26	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Horror', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
27	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sci-Fi/Fantasy', 'kind': 'youtube#videoCategory', 'etag': 'Ym...																																																																																																																																		
28	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Thriller', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
29	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shorts', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
30	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shows', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
31	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Trailers', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
Field	Annotations																																																																																																																																		
1	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Film & Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
2	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Autos & Vehicles', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
3	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Music', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...																																																																																																																																		
4	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Pets & Animals', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
5	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sports', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1BQ...																																																																																																																																		
6	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Short Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...																																																																																																																																		
7	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Travel & Events', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
8	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Gaming', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
9	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Videoblogging', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
10	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'People & Blogs', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
11	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
12	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Entertainment', 'kind': 'youtube#videoCategory', 'etag': 'Ym...																																																																																																																																		
13	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'News & Politics', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
14	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Howto & Style', 'kind': 'youtube#videoCategory', 'etag': 'Ym2...																																																																																																																																		
15	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Education', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
16	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Science & Technology', 'kind': 'youtube#videoCategory', 'eta...																																																																																																																																		
17	['assignable': true, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Nonprofits & Activism', 'kind': 'youtube#videoCategory', 'eta...																																																																																																																																		
18	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Movies', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
19	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Animation', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
20	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Classics', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
21	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Action/Adventure', 'kind': 'youtube#videoCategory', 'etag': 'Y...																																																																																																																																		
22	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Comedy', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
23	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Documentary', 'kind': 'youtube#videoCategory', 'etag': 'Ym...																																																																																																																																		
24	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Drama', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
25	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Family', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
26	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Foreign', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
27	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Horror', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
28	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Sci-Fi/Fantasy', 'kind': 'youtube#videoCategory', 'etag': 'Ym...																																																																																																																																		
29	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Thriller', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
30	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shorts', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
31	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Shows', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		
32	['assignable': false, 'channelId': 'UCBR8-60-828hp2BmDPantC', 'title': 'Trailers', 'kind': 'youtube#videoCategory', 'etag': 'Ym2ys1B...																																																																																																																																		

Slika 3: Upoređivanje kategorija

Obzirom da ne bismo imali sa čime da je poredimo, odlučili smo da sve podatke iz skupa USvideos sa tom kategorijom uklonimo. Prethodnom analizom skupova USVideos i GBVideos takođe je primećeno da se u skupu GBVideos nijednom ne pojavljuje video iz kategorije “Trailers” (category_id=43) pa su svi takvi uklonjeni iz skupa USVideos korišćenjem čvora Select.

	category_id
1	1
2	2
3	10
4	15
5	17
6	19
7	20
8	22
9	23
10	24
11	25
12	26
13	27
14	28
15	29

	category_id
1	1
2	2
3	10
4	15
5	17
6	19
7	20
8	22
9	23
10	24
11	25
12	26
13	27
14	28
15	29
16	43

Slika 4: Poređenje kategorija za skupove USvideos i GBvideos

Kako su oba Videos skupa činili podaci o video snimcima prikupljeni u toku nekoliko različitih dana, zarad izbegavanja redundantosti izvršene su redukcije. Naime, za svaki video klip u tabelama USVideos i GBVideos uzet je samo poslednji podatak i to korišćenjem čvora Aggregate (primarni ključ je samim tim samo video_id,a ne par video_id,date). Analogno su izdvojeni i jedinstveni komentari u skupovima USComments GBComments.

Nakon ovih izmena, spajane su tabele Videos i Comments i potom izvršeno grupisanje tabela po broju kategorije kako bi bile posmatrani prosek i maksimalna vrednost narednih karakteristika:

- Broj pregleda
- Broj lajkova
- Broj dislajkova
- Broj komentara
- Broj lajkova na komentaru
- Broj odgovora na komentar

Za broj pregleda je takođe posmatrana i najmanja vrednost za svaku kategoriju a statistike po kategorijama su prikazane na sledećim slikama.

GB_Statistics (14 fields, 14 records)

	category_id	views_Mean	views_Min	views_Max	likes_Mean	likes_Max	dislikes_Mean	dislikes_Max	comment_total_Mean	comment_total_Max	comment_likes_Mean	comment_likes_Max	replies_Mean	replies_Max
1	1	1834699.038	6945	6779570	36897.747	138606	1750.283	14402	4684.572	36254	2.852	23153	0.193	500
2	2	3188802.641	50487	11387085	23837.974	47761	1021.584	2248	2917.638	11455	2.063	3257	0.346	172
3	10	4455937.376	6451	58961407	159123.878	2289911	5778.321	143464	19187.650	813322	1.107	16830	0.107	504
4	15	1467062.095	6003	3611553	32564.004	71737	615.304	1918	4561.020	14898	0.877	840	0.064	40
5	17	1571603.433	2654	14960281	42487.953	446880	1099.375	9969	4326.367	43529	2.516	4740	0.468	308
6	19	298694.009	12323	474775	7388.086	11029	178.276	331	969.001	1853	1.210	160	0.156	62
7	20	858542.226	2116	3797928	30253.674	158760	1249.528	8481	5890.784	26137	1.469	3327	0.295	500
8	22	1486797.354	517	17521230	52622.310	437405	1975.190	41138	5661.871	72329	3.513	18019	0.255	501
9	23	2224578.236	2315	15314092	97853.235	969570	2489.500	35387	10528.800	180101	3.005	20860	0.175	501
10	24	2753559.786	2235	35300501	74941.102	1022015	5347.380	192725	10811.317	238974	2.824	23490	0.185	501
11	25	1259166.735	7786	10397787	8351.111	39633	1114.903	7524	2678.778	14426	1.193	2452	0.193	138
12	26	732599.458	15399	5551822	35997.236	505334	973.579	18490	4226.427	83992	4.489	60630	0.219	521
13	27	1072752.651	23850	3512713	35630.544	139562	1734.982	7494	4865.150	19943	3.234	4540	0.332	434
14	28	1300142.031	29752	8443106	28201.899	86006	1700.204	15504	3361.896	13233	0.883	2629	0.148	283

Slika 5: Statistika za podatke iz spojenih tabela GBvideos i GBcomments

US_Statistics (14 fields, 14 records)

	category_id	views_Mean	views_Min	views_Max	likes_Mean	likes_Max	dislikes_Mean	dislikes_Max	comment_total_Mean	comment_total_Max	comment_likes_Mean	comment_likes_Max	replies_Mean	replies_Max
1	1	1765711.872	2036	6694485	38036.751	163078	1966.207	15360	5644.281	36254	7.095	48772	0.267	501
2	2	1384832.237	3541	5564352	30805.924	116228	920.375	2607	3009.879	11130	0.949	439	0.326	89
3	10	2386918.814	0	41500672	106601.051	2010366	4341.007	135925	12157.491	736179	1.085	16830	0.105	504
4	15	1039376.918	3680	3611801	25473.167	71737	452.006	1918	3580.499	14898	1.126	1892	0.109	111
5	17	1541544.573	1097	13688054	36649.582	432783	1084.183	9367	4041.759	43546	1.664	1966	0.418	443
6	19	889119.379	19606	3369055	22404.916	64351	578.165	2358	2401.314	5468	1.578	1710	0.325	205
7	20	1472288.999	1597	5475859	39006.492	121972	2141.245	7322	8697.588	28192	2.086	2355	0.266	61
8	22	1718676.683	1341	15579127	66972.796	706104	6855.421	318404	9159.302	135649	3.810	17732	0.297	520
9	23	1762617.107	8886	13740234	80289.416	408649	2634.302	34393	8133.054	63436	3.752	20738	0.223	529
10	24	2089944.768	2143	33191594	60781.742	1075133	4351.828	184310	9364.936	337603	2.388	23295	0.193	500
11	25	859275.637	1731	10431296	7001.078	50802	1553.117	24306	2631.631	14679	1.448	2292	0.361	252
12	26	817833.996	8544	8059718	43583.885	656526	996.578	17328	5995.393	172790	4.351	27854	0.288	507
13	27	918187.250	2482	5502137	28087.735	137078	1342.629	13695	3892.991	21707	2.520	4541	0.328	409
14	28	1174408.984	2845	4804702	30746.542	146049	1926.284	29293	4387.230	22556	2.303	16760	0.303	504

Slika 6: Statistika za podatke iz spojenih tabela USvideos i UScomments

Uočavamo da snimci skupljeni u Velikoj Britaniji imaju veći prosečan broj pregleda u čak 10/14 analiziranih kategorija. Ipak, ova statistika je jedina sa značajnom razlikom u odnosu. Naime, kada su u pitanju prosečan broj lajkova i komentara imamo podjednak broj dominantnih kategorija za obe regije. Takođe, prosečan broj dislajkova za američke snimke veći od onog za britanske u 8/14 slučajeva. Zanimljivo je da neke kategorije u određenoj regiji imaju veće vrednosti svakog parametra u odnosu na vrednosti iste kategorije u drugoj regiji. Primeri ovoga jesu za Veliku Britaniju kategorije „Comedy“ i „Anime/Animation“ (oznake kategorija redom 10 i 17) i za SAD kategorija „Classics“ sa oznakom 19. Takođe vidimo da u USVideos postoji jedan video u kategoriji 10 koji nema nijedan pregled, što bi ukazivalo na pogrešan unos jer se u spojenim tabelama nalaze samo oni video snimci koji imaju komentare, a nije moguće ostaviti komentar bez toga da se ulazak na video registruje.

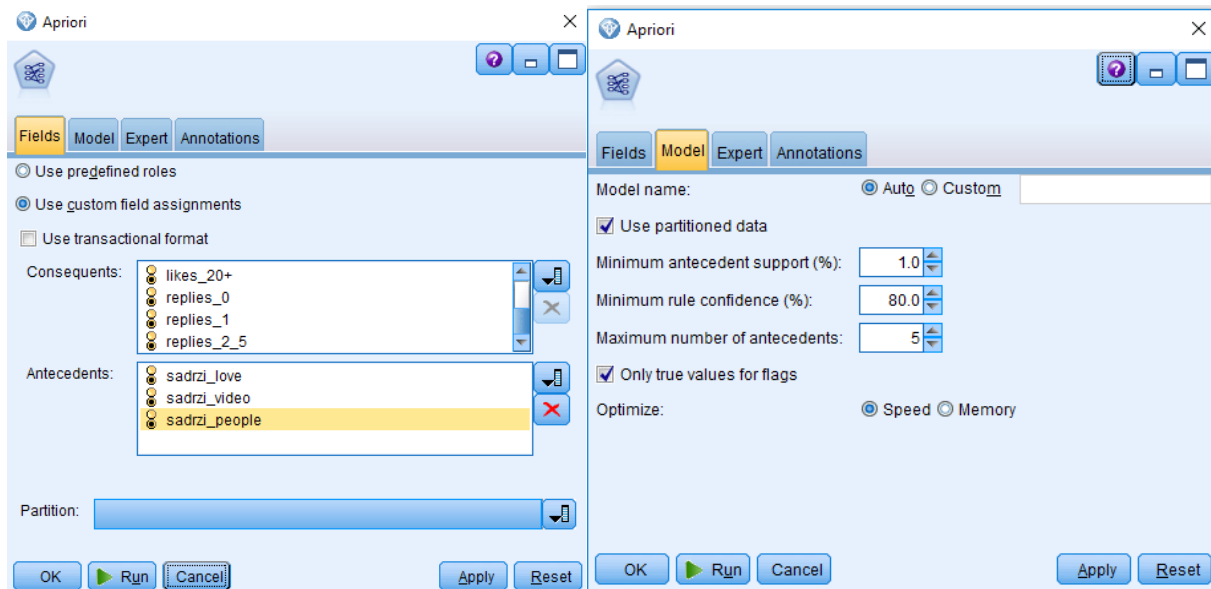
Pravila pridruživanja

Različitim algoritmima pravila pridruživanja pokušali smo da uočimo neke zavisnosti podataka. Prva ideja je bila da proverimo kako se komentari koji sadrže neke od najčešćih reči odnose po broju lajkova i odgovora. Kako bismo dobili reči koje se najčešće pojavljuju u komentarima koristili smo alat pod nazivom Text Analytics. Za skup podataka USComments rezultat je sledeći:



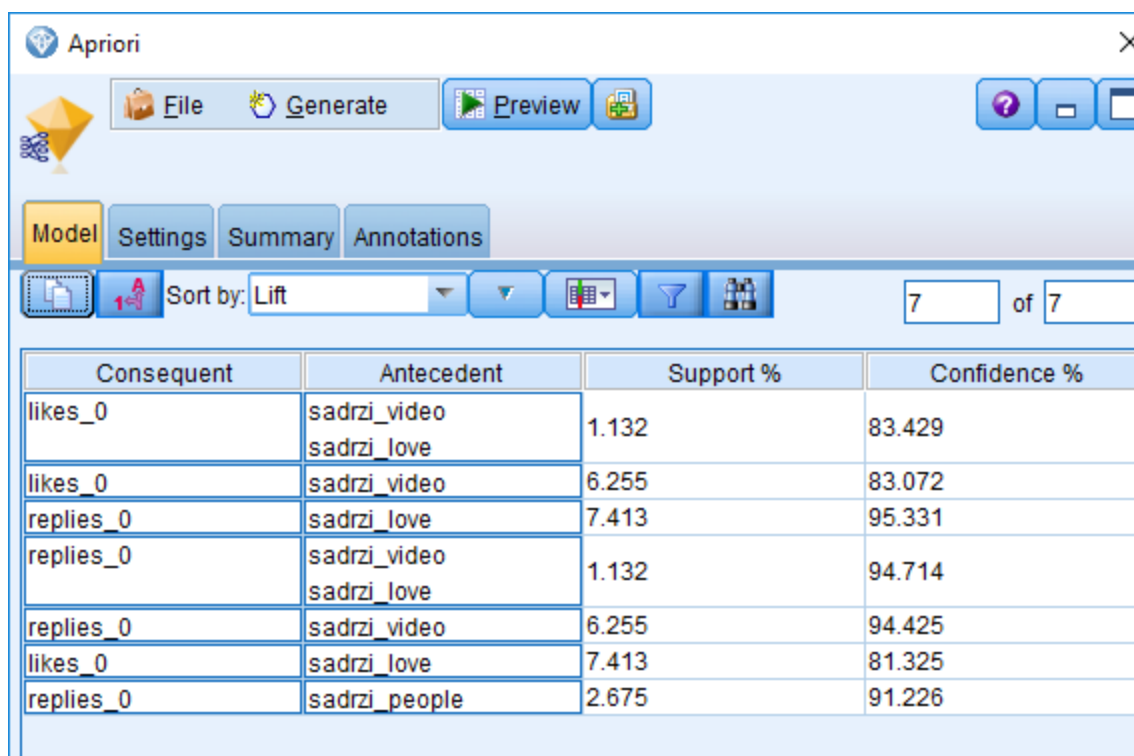
Slika 9: Formatiranje podataka za algoritme

Svaki komentar je sada imao nove atribute tzv. flegove koji su govorili da li sadrži određenu reč, i označavali pripadnost nekoj od klasa za broj lajkova i odgovora. Komentari su po broju lajkova podeljeni u pet klasa, dok su za broj odgovora podeljeni u četiri klase. Pomenuta kategorizacija za broj komentara i odgovora izvršena je pravljenjem novih kolona korišćenjem čvorova Derive. Primenjen je ovaj pristup jer se korišćenjem čvora Binning većina snimaka klasifikuje u jednu klasu („najmanju“ po oba parametra) bez obzira na odabir raspodele. Nakon ovoga, podaci su bili u odgovarajućem formatu za korišćenje Apriori algoritma, čiji su parametri bili podešeni na sledeći način.



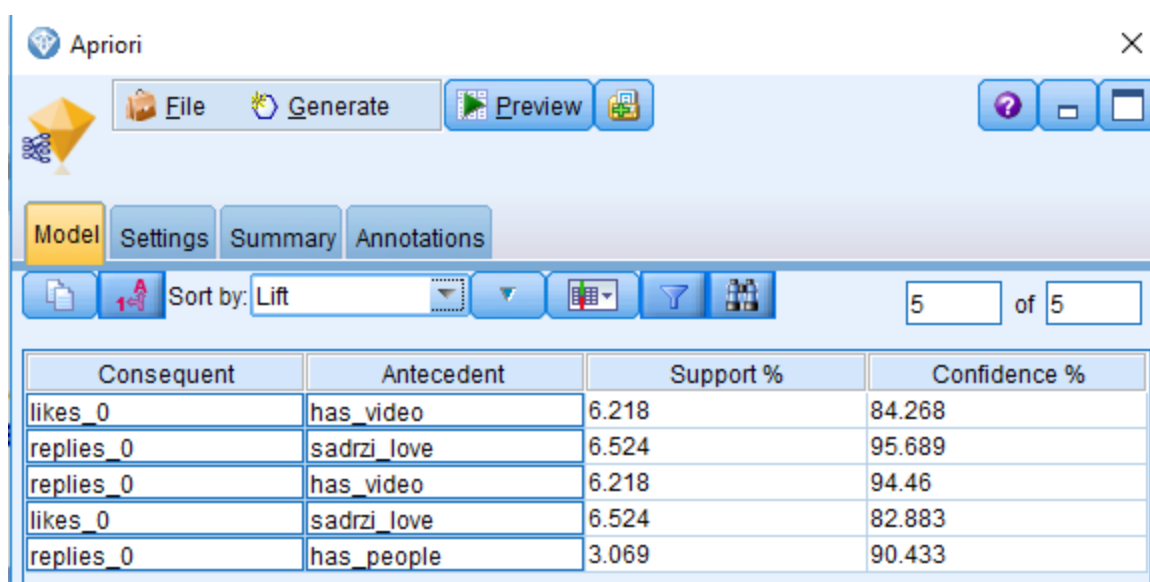
Slika 10: Ulazni podaci za Apriori algoritam Slika 11: Parametri za Apriori algoritam

Dobijeni rezultati za skupove GBComments i USComments nalaze se redom na sledećim slikama:



Consequent	Antecedent	Support %	Confidence %
likes_0	sadrzi_video sadrzi_love	1.132	83.429
likes_0	sadrzi_video	6.255	83.072
replies_0	sadrzi_love	7.413	95.331
replies_0	sadrzi_video sadrzi_love	1.132	94.714
replies_0	sadrzi_video	6.255	94.425
likes_0	sadrzi_love	7.413	81.325
replies_0	sadrzi_people	2.675	91.226

Slika 12: Rezultati za skup GBcomments primenom Apriori algoritma



Consequent	Antecedent	Support %	Confidence %
likes_0	has_video	6.218	84.268
replies_0	sadrzi_love	6.524	95.689
replies_0	has_video	6.218	94.46
likes_0	sadrzi_love	6.524	82.883
replies_0	has_people	3.069	90.433

Slika 13: Rezultati za skup USComments primenom Apriori algoritma

Dobijena pravila su poprilično slična, što i ne čudi imajući u vidu da najveći deo komentara spada u „najnižu“ kategoriju po oba parametra (broj lajkova/odgovora). Da ova pravila nisu previše zanimljiva govori i lift mera koja je za sva pravila u opsegu 0.962% - 1.016% za GBcomments, i 0.954% - 1.013% za USComments.

Sa druge strane, primenom Carma algoritma koji koristi sva polja koje je koristio i Apriori algoritam i ima iste parametre dobijeni su sledeći rezultati redom za GBComments i USComments.

Sort by: Lift				17	of 17
Consequent	Antecedent	Support %	Confidence %		
likes_0	sadrzi_video replies_0	5.906	85.192		
replies_0	sadrzi_love likes_0	6.029	97.449		
likes_0	replies_0	94.843	84.096		
replies_0	likes_0	82.107	97.141		
replies_0	sadrzi_video likes_0	5.196	96.835		
likes_0	sadrzi_love replies_0	7.067	83.132		
likes_0	sadrzi_video	6.255	83.072		
likes_0 replies_0	sadrzi_video	6.255	80.443		
likes_0	sadrzi_people replies_0	2.44	82.726		
replies_0	sadrzi_love	7.413	95.331		
replies_0	sadrzi_people likes_0	2.121	95.167		
replies_0	sadrzi_love sadrzi_video	1.132	94.714		
replies_0	sadrzi_video	6.255	94.425		
likes_0	sadrzi_love	7.413	81.325		
replies_0	likes_1	11.219	92.735		
replies_0	sadrzi_people	2.675	91.226		
replies_0	likes_2_5	4.534	82.08		

Slika 14: Rezultati za skup GBcomments primenom Carma algoritma

Consequent	Antecedent	Support %	Confidence %		
replies_0	sadrzi_love likes_0	5.407	97.759		
replies_0	likes_0	83.226	97.181		
replies_0	has_video likes_0	5.24	96.84		
replies_0	sadrzi_love	6.524	95.689		
replies_0	has_people likes_0	2.442	94.529		
replies_0	has_video	6.218	94.46		
replies_0	likes_1	10.843	92.36		
replies_0	has_people	3.069	90.433		
likes_0	has_video replies_0	5.873	86.391		
likes_0	replies_0	94.832	85.288		
likes_0	sadrzi_love replies_0	6.242	84.675		
likes_0	has_video	6.218	84.268		
likes_0	has_people replies_0	2.776	83.176		
likes_0	sadrzi_love	6.524	82.883		
likes_0 replies_0	has_video	6.218	81.605		
likes_0 replies_0	sadrzi_love	6.524	81.025		

Slika 15: Rezultat za skup UScomments primenom Carma algoritma

Ponovo, pravila nisu preterano zanimljiva i međusobno su veoma slična, i sva za posledicu imaju klasifikaciju snimaka u najnižu kategoriju po lajkovima i odgovorima. Poređenja radi Lift mera je za ova pravila u opsegu 0.675%-1.038% za GBcomments i 0.954%-1.038% za UScomments.

Sledeća ideja je bila da izdvojimo tagove koji se najčešće pojavljuju u snimcima, i da proverimo da li se, i koji od njih pojavljuju zajedno. Opet smo korišćenjem Text Analytics alata analizirali atribut tags u oba skupa podataka- USvideos i GBvideos. Rezultat je redom:

14,802 concepts

	Concept	In	Global	Docs	Type
1	iphone		1254	155 (2%)	<Product>
2	comedy		569	516 (6%)	<Unknown>
3	news		563	396 (5%)	<Unknown>
4	show		529	266 (3%)	<Unknown>
5	makeup		451	333 (4%)	<Unknown>
6	review		423	311 (4%)	<Unknown>
7	video		419	370 (5%)	<Unknown>
8	trailer		384	324 (4%)	<Unknown>
9	interview		381	290 (4%)	<Unknown>
10	celebrity		372	312 (4%)	<Unknown>
11	movie		357	292 (4%)	<Unknown>
12	food		327	296 (4%)	<Unknown>

1 / 2

Slika 16: Najčešći tagovi u skupu USvideos

11,798 concepts

	Concept	In	Global	Docs	Type
1	vlog		765	452 (6%)	<Unknown>
2	iphone		623	100 (1%)	<Product>
3	makeup		556	407 (5%)	<Unknown>
4	interview		479	347 (4%)	<Unknown>
5	daily		470	371 (5%)	<TimePeriod>
6	comedy		432	425 (5%)	<Unknown>
7	news		414	257 (3%)	<Unknown>
8	show		394	207 (3%)	<Unknown>
9	review		390	336 (4%)	<Unknown>
10	video		382	366 (5%)	<Unknown>
11	movie		359	263 (3%)	<Unknown>
12	fashion		351	337 (4%)	<Unknown>

1 / 2

Slika 17: Najčešći tagovi u skupu GBvideos

Vidimo da su na prvih pet pozicija na snimcima iz SAD-a to: iphone, comedy, news, show i makeup. A na snimcima iz Velike Britanije: vlog, iphone, makeup, interview i daily. Napravili smo novu kolonu za svaki od tagova koja govori da li određeni video sadrži taj tag ili ne.

Primenom algoritma Apriori i Carma(slika 18) na jedinstvene video snimke iz skupa podataka GBvideos dobijamo sledeći rezultat:

Apriori Results (Left):

Consequent	Antecedent	Support %	Confidence %
sadrzi_vlog	sadrzi_daily	3.896	84.058

Carma Results (Right):

Consequent	Antecedent	Support %	Confidence %
sadrzi_vlog	sadrzi_daily	18.598	84.058

Slika 18: Rezltati za algoritam Apriori(levo) i Carma(desno) na skupu podataka GBvideos

U oba slučaja dobijamo jedno pravilo, i to je da se tag vlog pojavljuje zajedno sa tagom daily. Rekli bismo da ovo ima smisla, obzirom da se na ovim vrstama blogova (tzv. vlog) klipovi izbacuju često, možda i svakodnevno (eng. Daily).

Razlika koju uočavamo je da za algoritam Apriori dobijamo pravilo sa dosta manjom podrškom nego što je slučaj za algoritam Carma. Da smo za oba algoritma koristili istu podršku, na primer 5%, dobili bismo rešenja samo za algoritam Carma. Sada uzimamo drugi skup tagova koji se odnosi na klipove iz SAD-a:

Apriori Results (Left):

Consequent	Antecedent	Support %	Confidence %
sadrzi_show	sadrzi_comedy	8.24	35.678
sadrzi_comedy	sadrzi_show	8.778	33.491

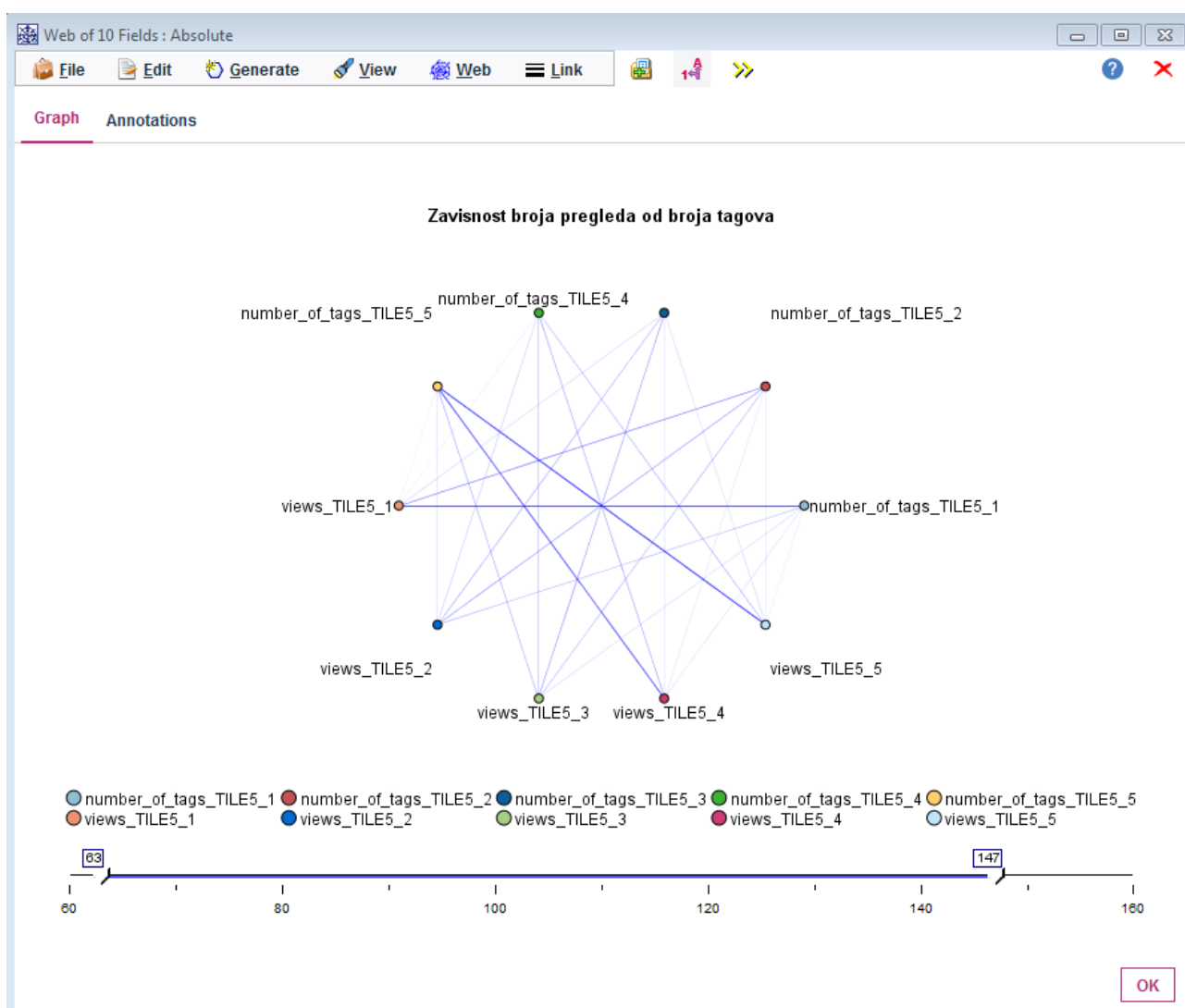
Carma Results (Right):

Consequent	Antecedent	Support %	Confidence %
sadrzi_show	sadrzi_comedy	32.677	35.678
sadrzi_comedy	sadrzi_show	34.811	33.491

Slika 19: Rezultati za algoritam Apriori(levo) i Carma(desno) na skupu podatka USvideos

U ovom slučaju dobijamo po dva ista pravila u oba algoritma, a to su da uz tag show obično ide i tag comedy, i obrnuto.

Zatim nas je zanimalo kako broj tagova na određenom snimku utiče na broj pregleda istog, odnosno da li bi veći broj tagova povlačio i veću popularnost. Za svaku instancu smo izdvojili broj tagova koje sadži. Te brojeve smo zatim podelili u 5 kategorija. Takođe smo broj pregleda podelili u 5 kategorija. Za svaku od tih kategorija smo dodali po nov atribut koji ima vrednost tačno, odnosno netačno ako snimak pripada, odnosno ne pripada toj kategoriji. Hajde prvo da pogledamo zavisnost novonapravljenih kategorija preko vizuelnog prikaza podataka. Koristićemo grafik koji prikazuje podatke u okviru mreže. Svaki čvor mreže je jedna kategorija. Intenzitet linije koja spaja čvorove nam ukazuje na intenzitet povezanosti kategorija.

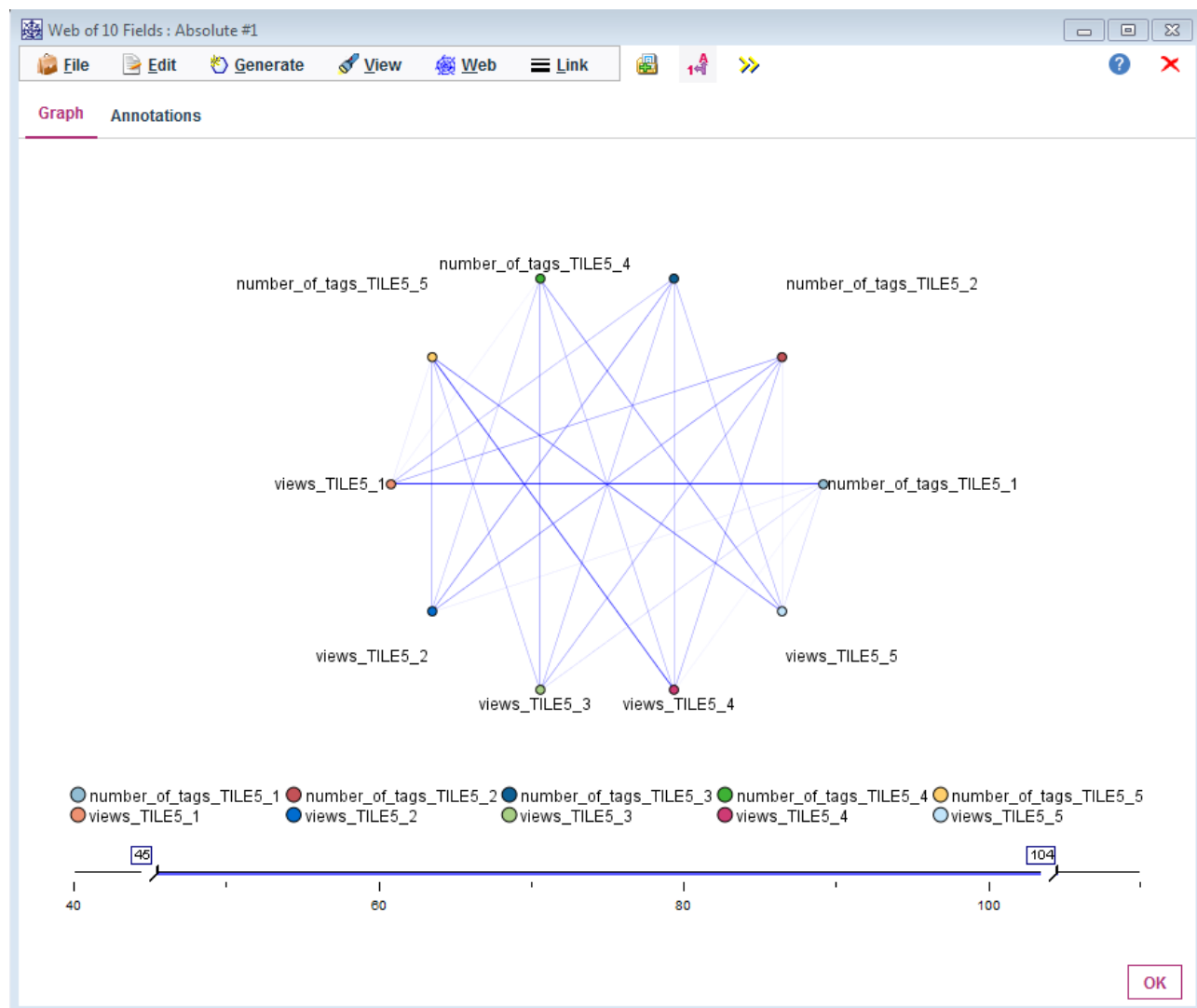


Slika 20: Grafik zavisnosti kategorija u skupu USvideos

Na slici iznad (Slika 20) je prikaz zavisnosti kategorija u okviru podataka USvideos. Veći broj kategorije znači veći broj pregleda, odnosno tagova. Analogno važi za manji broj. Na grafiku imamo tri linije jačeg intenziteta, te nas zanima koje veze su u pitanju. Vidimo da je kategorija 5 broja tagova povezana sa kategorijom 5, i kategorijom 4 broja pregleda. To ukazuje da je veći broj

tagova povezan sa većim brojem pregleda. Takođe vidimo da je kategorija 1 broja tagova povezana sa kategorijom 1 broja pregleda. Odavde bi moglo da važi da mali broj tagova znači i manji broj pregleda.

Pogledajmo grafički prikaz povezanosti kategorija i na skupu GBvideos.



Slika 21: Grafik zavisnosti kategorija u skupu GBvideos

Što se tiče podataka vezanih za snimke u GBvideos situacija je slična, sa izuzetkom da najveći broj tagova nije intenzivno povezan sa najvećim brojem pregleda. Naime, imamo povezanost između kategorije 5 broja tagova i kategorije 4 broja pregleda. Važi povezanost između malog broja tagova i pregleda.

Proverićemo ove zavisnosti primenom algoritama za pravila pridruživanja. Za oba skupa podataka i oba algoritma parametri su podešeni na isti način. Korišćenjem algoritma Apriori za skupove USvideos(levo) i GBvideos(desno) dobijamo sledeći rezultat:

Consequent	Antecedent	Support %	Confidence %
views_TILE5_1	number_of_tags_TILE5_1	19.789	30.769
views_TILE5_5	number_of_tags_TILE5_5	22.241	27.757
views_TILE5_4	number_of_tags_TILE5_5	22.241	25.856

Consequent	Antecedent	Support %	Confidence %
views_TILE5_1	number_of_tags_TILE5_1	18.203	32.595
views_TILE5_4	number_of_tags_TILE5_5	21.717	25.464

Slika 22: Rezultat primene Apriori algoritma na podatke iz skupa USvideos(levo) i GBvideos(desno)

Algoritam Apriori potvrđuje pretpostavke izvedene na osnovu grafika. Ulaz za algoritam je broj tagova, a cilj broj pregleda. Što je veći broj tagova veći je i broj pregleda. A ako imamo mali broj tagova, šanse su da će biti i mali broj pregleda.

Pogledajmo rezultat dobijen algoritmom Carma:

Consequent	Antecedent	Support %	Confidence %
number_of_tags_TILE5_5	views_TILE5_5	20.0	30.867
views_TILE5_1	number_of_tags_TILE5_1	19.789	30.769
number_of_tags_TILE5_1	views_TILE5_1	20.0	30.444
number_of_tags_TILE5_5	views_TILE5_4	20.0	28.753
views_TILE5_5	number_of_tags_TILE5_5	22.241	27.757
views_TILE5_4	number_of_tags_TILE5_5	22.241	25.856

Consequent	Antecedent	Support %	Confidence %
views_TILE5_1	number_of_tags_TILE5_1	18.203	32.595
number_of_tags_TILE5_1	views_TILE5_1	20.046	29.598
number_of_tags_TILE5_5	views_TILE5_4	19.988	27.666
views_TILE5_4	number_of_tags_TILE5_5	21.717	25.464

Slika 23: Rezultat primene Carma algoritma na podatke iz skupa USvideos(levo) i GBvideos(desno)

Algoritam Carma daje iste rezultate, s tim da ovde možemo videti i kako broj pregleda utiče na broj tagova, ali to nije nešto što je previše smisleno za istraživanje, pa ćemo se fokusirati na pravila koja govore kako broj tagova utiče na broj pregleda. I u ovom slučaju su pretpostavke izvedene iz grafika povezanosti potvrđene.