Home   Blog   Jobs   Trainings   Learning Paths

• j

Analytics Vidhya
Learn Everything About Analytics

👤 Log In

📌 Predict the gem of Auxesia for Magazino!

Hackathons   Online hackathon

---

kunal 🛡 Data Hackers

1 ✏ 1h

At Analytics Vidhya, we love creating new content. That is the way we have grown!

While we have relied mostly on listening to our audience and simplifying what is difficult for them, how about putting little science to this? That is exactly what we will be doing today for our friend Magazino!

Brief about Magazino

Magazino, was started by a few technoprenuers about a year back, who wanted to bring data science to publishing. They believed that current media houses are not doing justice to what should be served to people and with use of Big Data they can get a fair differentiation in their offering.

In the last one year, Magazino has done a lot of experimentation with publishing several topics and content. They have made sure that they capture all the data at the back end. Now, they want to put a bit of science to predict which articles get shared on social media, which ones don't! To that extent, they have stored social media performance of ~24,875 articles they have published till date along various attributes.

What is the gem of Auxesia?

Auxesia is the goddess of growth and prosperity in Greek mythology! Finding articles which can go viral in today's age is like being blessed with a gem from Auxesia. So, you have to find out these gems for Magazino!

Description of the dataset:

Here is a brief description of the attributes:
1. **id** - Unique id of the article
2. **n_tokens_title:** Number of words in the title
3. **n_tokens_content:** Number of words in the content
4. **n_unique_tokens:** Rate of unique words in the content
5. **n_non_stop_words:** Rate of non-stop words in the content
6. **n_non_stop_unique_tokens:** Rate of unique non-stop words in the content
7. **num_hrefs:** Number of links
8. **num_self_hrefs:** Number of links to other articles published by Magazino
9. **num_imgs:** Number of images
10. **num_videos:** Number of videos
11. **average_token_length:** Average length of the words in the content
12. **num_keywords:** Number of keywords in the metadata
13. **Category_article** - Various categories of article
14. **kw_min_min:** Worst keyword (min. shares)
15. **kw_max_min:** Worst keyword (max. shares)
16. **kw_avg_min:** Worst keyword (avg. shares)
17. **kw_min_max:** Best keyword (min. shares)
18. **kw_max_max:** Best keyword (max. shares)

1 / 2 ⬍

19. **kw_avg_max:** Best keyword (avg. shares)
20. **kw_min_avg:** Avg. keyword (min. shares)
21. **kw_max_avg:** Avg. keyword (max. shares)
22. **kw_avg_avg:** Avg. keyword (avg. shares)
23. **self_reference_min_shares:** Min. shares of referenced articles in Magazino
24. **self_reference_max_shares:** Max. shares of referenced articles in Magazino
25. **self_reference_avg_shares:** Avg. shares of referenced articles in Magazino
26. **Day_of_publishing:** Day of week - Monday, Tuesday.......Sunday
27. **LDA_00:** Closeness to LDA topic 0
28. **LDA_01:** Closeness to LDA topic 1
29. **LDA_02:** Closeness to LDA topic 2
30. **LDA_03:** Closeness to LDA topic 3
31. **LDA_04:** Closeness to LDA topic 4
32. **global_subjectivity:** Text subjectivity
33. **global_sentiment_polarity:** Text sentiment polarity
34. **global_rate_positive_words:** Rate of positive words in the content
35. **global_rate_negative_words:** Rate of negative words in the content
36. **rate_positive_words:** Rate of positive words among non-neutral tokens
37. **rate_negative_words:** Rate of negative words among non-neutral tokens
38. **avg_positive_polarity:** Avg. polarity of positive words
39. **min_positive_polarity:** Min. polarity of positive words
40. **max_positive_polarity:** Max. polarity of positive words
41. **avg_negative_polarity:** Avg. polarity of negative words
42. **min_negative_polarity:** Min. polarity of negative words
43. **max_negative_polarity:** Max. polarity of negative words
44. **title_subjectivity:** Title subjectivity
45. **title_sentiment_polarity:** Title polarity
46. **abs_title_subjectivity:** Absolute subjectivity level
47. **abs_title_sentiment_polarity:** Absolute polarity level
48. **shares:** Number of shares (target)

Where polarity and subjectivity are measured on a relative scale. All the remaining terminology are standard in any publishing world. You have title of articles and keywords associated with the articles. If you are not sure about something, just google it!

Objective

They also have more than 10,000 articles in pipeline and they want to predict which of these are piping hot to share on social media! So, here you go, train dataset contains the information of all these 24,875 articles while test contains the 10,089 articles in the pipeline. You need to predict the number of shares for each article in the pipeline.

Dataset:

⬇ sample_submission.csv.zip (11.1 KB) ⬇ test.csv.zip (1.1 MB) ⬇ train.csv.zip (2.8 MB)

P.S. If you are not able to download the dataset

Evaluation:

The evaluating criteria for this problem would be Root Mean Square error (RMSE). We will be putting up a solution checker online on Saturday afternoon, where you can upload your solutions and check their accuracy based on a small sample. The URL of the same will be released tomorrow.

Expected code of conduct:

- The aim of this hackathon is to create an environment of collective learning in Analytics Vidhya community. If you are facing a conflict at any point in time, remember this and take the

1 / 2

- If you are facing any challenge where you need help on, please use the discussion portal. Keep the discussions specific and useful. If they are not, they will be deleted. You should post these discussions in the category - Hackathons - Online Hackathon
- For non-technical discussions and saying high-fives to the community, use the slack channel. People are already rolling up their sleeves there. You should have got an invite to the slack channel for discussions already. If not, please drop me a PM - you should join the fun there.
http://analyticsvidhya.slack.com 9
- You are free to use whatever software you wish, as long as you have rightful access to it
- The winner takes away Amazon vouchers worth INR 10,000 (~$200). You would need to share your solution and the code with the community.
- We are learning as much about the community building as you are about data science. So, if you see a bug (yes, those creepy creatures!), please raise it with us through slack or PM any member of the team directly
- At the end of the contest, we will share the larger dataset and the correct answers for the problem.

At the end of the day - have fun, help your fellow members and hope to find out gems of Auxesia for Magazino!

3 people liked this.

created	last reply	1	124	1	3	5	
🧑 1 hour	🧑 1 hour	reply	views	user	likes	links	⌄

---

kunal 🛡 Data Hackers   1h

@Srikanth_Adepu , @Atul_Sharma , @Steve , @karthe1 , @dharmesh_manglani029 , @Ramesh_Ramachandran , @sreeramindia , @saimadhup , @sarthak93 , @ramanindya55 , @anon , @vajravi , @gauravkantgoel , @dkanand86 , @Amaresh_Murthiraju , @abhishek_krishna , @nknithinabc , @santu_rcc014 , @Ishan , @Jitesh_Khandelwal , @Subarna , @kunal , @ParindDhillon , @adityashrm21 , @ParulChoudhary , @raghava_r4u , @abimannan , @Avioboy , @yh18190 , @Venkata , @Neeraj_Kumar , @numb3r303 , @SRK , @ashu , @krishna_cse37 , @sekhar_chandra630 , @Uday_Bhan_Singh , @kesavkulkarni , @san_sri12 , @Sumangla , @rithwik , @vikash , @Mounika , @Shyam_Naren , @praveen766 , @santoshgsk , @Sajan_Kedia , @nayan1247 , @aatishk , @Mohit_Bansal , @dhuwalia_sumit , @Just_Rahul , @sreak1089 , @saiki0684 , @metalosaur , @satya23k , @jitin_kapila , @Hankarthu , @singhrobin1302 , @Sarangam , @kanigalupula , @Basu , @manizoya_1 , @yadhu621 , @Vignesh , @dada_kishore , @sandeepgupta2 , @gaurav_ism , @rajendra_belwal , @pavitrakumar , @mrsan22 , @Veldandi_Karthik_Kum , @pradeep_1209 , @Nalin , @Rishabh0709 , @anshulkgupta93_1 , @debarati_dutta8 , @prateek_gs , @architv07 , @manojprabhakar9090 , @karthikdoesit , @bvdrao , @karthiv , @aliasZero , @shaz1985 , @promila , @debapriya28 , @vikashsinghy2k , @zs_master , @VIKAS12APRIL , @0_rishabh , @Debanjan_Banerjee , @Amber_Saxena , @ayesha20 , @Aksgupta123_1 , @sahil030288 , @ankitbhargava06 , @atul0017 , @pooja10 , @Sagar_Pardeshi , @Tushar_Kakkar , @mail1 , @Orca , @Vivek_Agarwal , @Deepak_Singhal , @softwarmechanic , @gauravkumar37 , @nathkundan , @dalvi2u , @Sai_Krishna_Bala , @dipanjandeb79 , @vishnuvardhanreddy40 , @mat04460 , @savioseb_1 , @GaurKaps , @Bhargavi_Singaraiah , @Dev_Anand , @Abinaya , @sumitgupta07 , @Debanjan_Chaudhuri , @vijayavel , @Anilkumar_Panda , @Akshay_Tiwari , @anuragbaid , @devanshrising , @Rupert , @Puneet_Rathore , @Jegan_Venkatasamy , @kparitoshprasad , @sumit , @G9hardik , @Prabakaran , @yash05 , @uditsaini88 , @Anurag , @lmsy , @punardeep , @Sitaram , @AbhishekSingh , @Kushagra_Sharma , @Ram_Marthi , @sidhusmart , @seaem , @amarjeetpatra , @JohnFelix , @Arunkumar_t , @katakwar_shashank , @Priyabrata_Dash , @Mvijay , @abhi , @sorabhkalra , @bhavyaghai , @Kaushik_Roy_Chowdhur , @animeesh2007asansol , @Sidhraj , @Amol_Pol , @akbism , @Karthigeyan_Thiyagar , @Sarah_Masud , @Prerit_Khandelwal , @raghavyadav990 , @badrinath_mn , @Santhosh_Kumar , @chidambara_natarajan @varun_khandelwal08 , @Shiva_Akella , @kirubakumaresh , @explorer , @sh

1 / 2 ⇕

@Mohd_Junaid , @ABHISHEK_SINGLA , @Apoorv_Anand , @amogh_kaliwal , @Risoni , @Anurag_GV , @Rajiv_Iyer , @mahanteshimath , @Rashmi1311 , @preethyvarma , @kripton99 , @litankumar , @Dhruv_Nigam , @santhudr , @Shubham_Jain , @Swapnil_Sharma , @Deepesh87 , @bala_io , @Ankit_Samantara , @barnava7 , @Ajay_Kumawat , @shrawankumarhari , @Number_Cruncher , @MJlnen , @umeshnrao , @KP1 , @anujp , @Prarthana_Bhat , @crptick , @Shobhit_Mittal_de_Ro , @Naseer , @NM8185 , @Malav_Shah , @Ved_Klawat , @poornaramakrishnan , @Abhishek_Nagarjuna , @Neel_Biswas , @amrrs , @Mohammed

  Log In to Reply

Suggested Topics

Topic	Category	Replies	Views	Activity
Hackathon Problem: Do you know who's a Megastar?	Online hackathon	75	1.7k	25d
Data Hackathon - Online (Date: 7th June 2015)	Online hackathon	144	2.7k	25d
Rules for the online hackathon	Online hackathon	6	237	Jun 6
Hackathon resource guide for tomorrow	Online hackathon	4	179	Jun 7
📌 Data Hackathon Online (11th & 12th July 2015) - Register Here	Online hackathon	217	4.6k	12m

Want to read more? Browse other topics in   **Online hackathon** or **view latest topics**.

1 / 2 ⇕