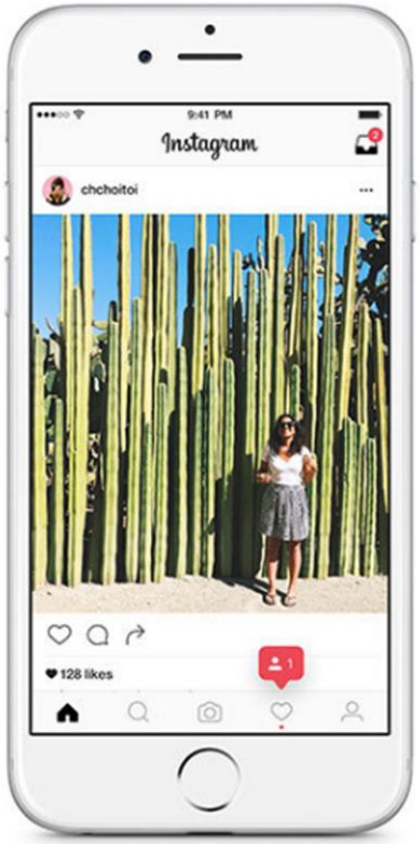
Predicting the Popularity of Instagram Posts for a Lifestyle Magazine Using Deep Learning

Shaunak De, Abhishek Maity, Vritti Goel, Sanjay Shitole and Avik Bhattacharya

Introduction

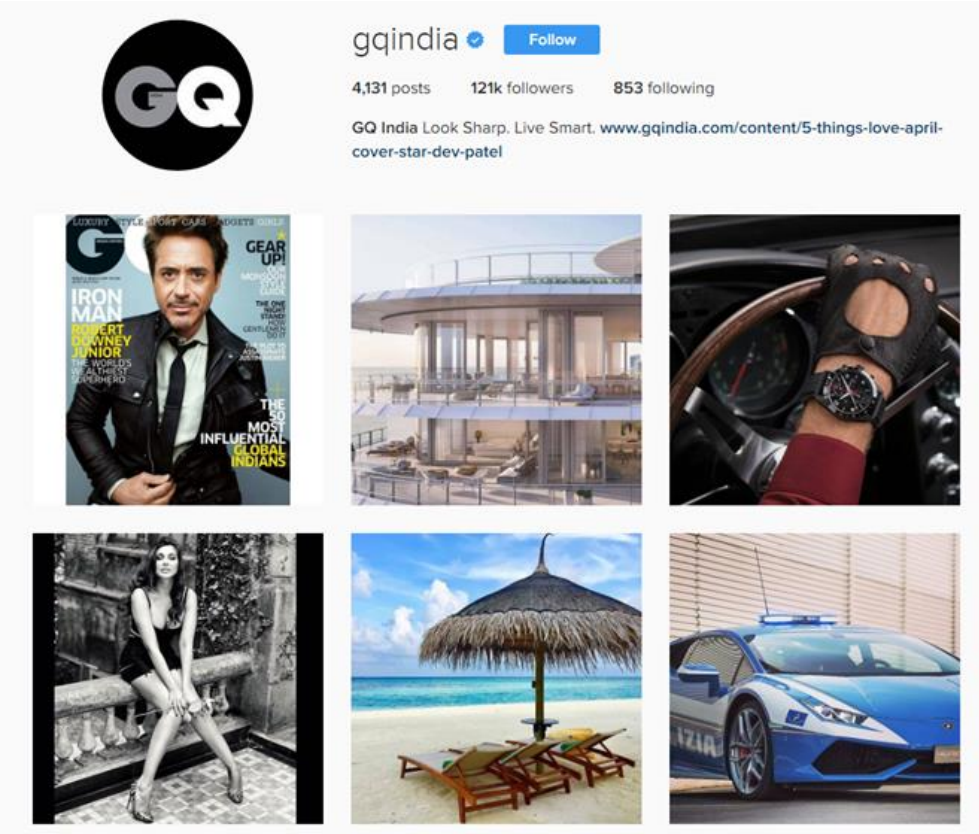
- **Instagram** is a social media platform for visual media-sharing.
- Increasingly being adopted by traditional media platforms like magazines.
- Analysis of the **popularity** or **traction** of the Instagram posts becomes important for estimation of reach etc.
- In a commercial scenario it is important to be able to coarsely **predict** the **reach** of a particular post for price fixation with advertisers.



"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Commercial Interest

- **Reach Analysis**
  - Estimation of reader interaction
  - Influence of magazine
  - Brand Image
- **Advertorial Price Fixing**
  - Pricing depends on reader interaction
  - Enforceable and measurable impact



"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Typical Post

"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Metrics

Feature Name	Sample Data Point
filterApplied	'Gingham'
creationTime	1482831357
weekOfTheYear	24
dayOfTheWeek	03
hourOfTheDay	06
URL	https://scontent-sin6-1.cdninstagram.com/t51.2885-15/e35/15625548_..._332762100315193344O_n.jpg
NumberTagged	5
Caption	The customary #polaroid. Photographer Tarun Vishwa goes #oldschool for beauty #KanganaRanaut during #GQAwards shoots. #BTS #Exclusive #Throwback #WomenWeLove #2016 #Woman
lengthCaption	174
numberOfTags	10
tagList	#polaroid #oldschool #KanganaRanaut #GQAwards #BTS #Exclusive #Throwback #WomenWeLove #2016 #Woman

"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Challenge of "Tag" discovery

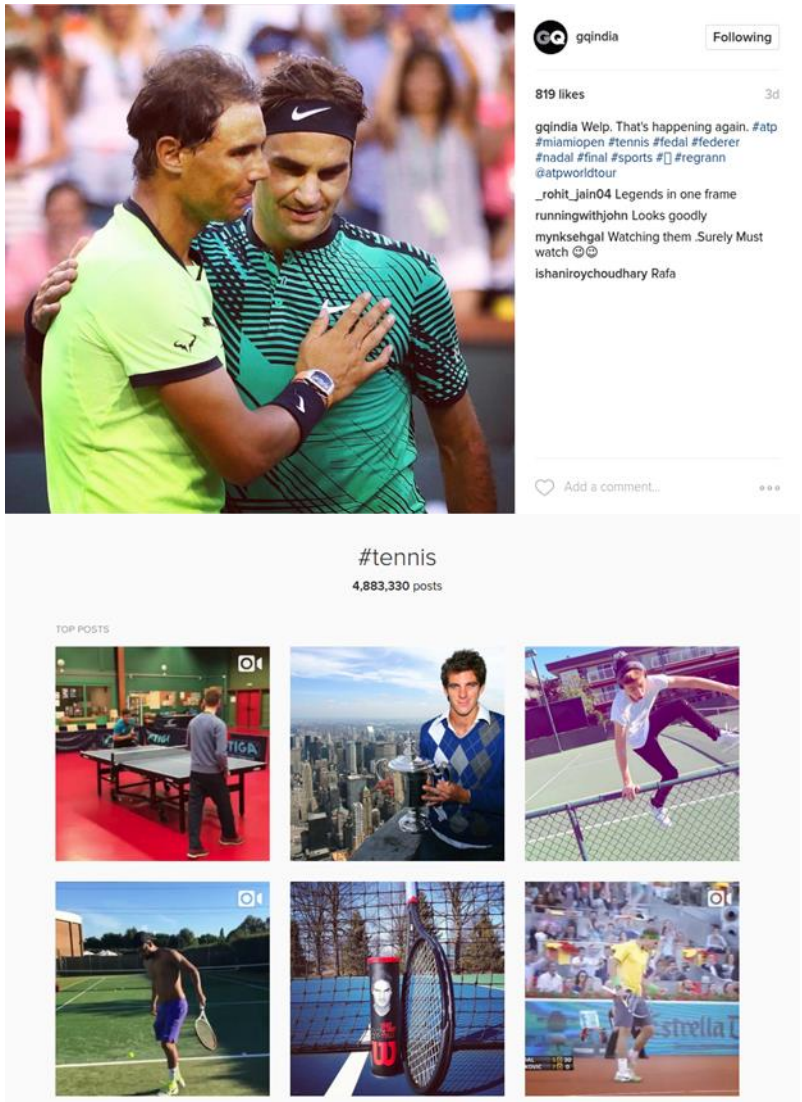
- #watches is related to a post tagged #seiko
- Because ‘Seiko‘ is a manufacturer of ‘watches’.
- However, lexicographically they have little inter-relation
- **Solution:**
  The use of a **word-tree**
  - Post **A** contains the tags #**watch**, #**cricket** and #**sachin**
  - Post **B** contains the tags #**cricket**, #**game**,
  - Then **both** posts are to be **grouped** into the same category.

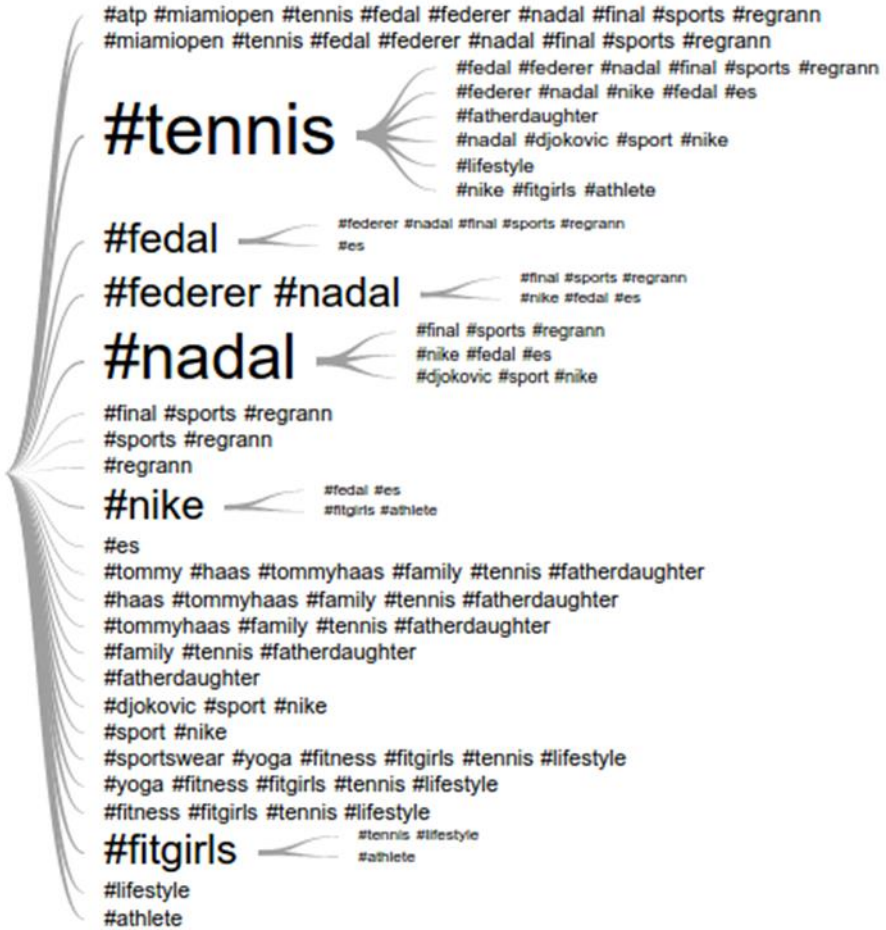
"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Challenge of the Word-Tree

- In practice however, this approach caused the grouping of a large number of unrelated posts
- Because certain **common tags** are repeated.
- **Solution: Pruning**
  - Ranked the tags by their **occurrence** and **deleted** 10% of the most **commonly used tags**.
  - This leads to **reasonable separation** of post categories.
  - Each tag category was **encoded** with a positive whole **number** and applied to the posts.

"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

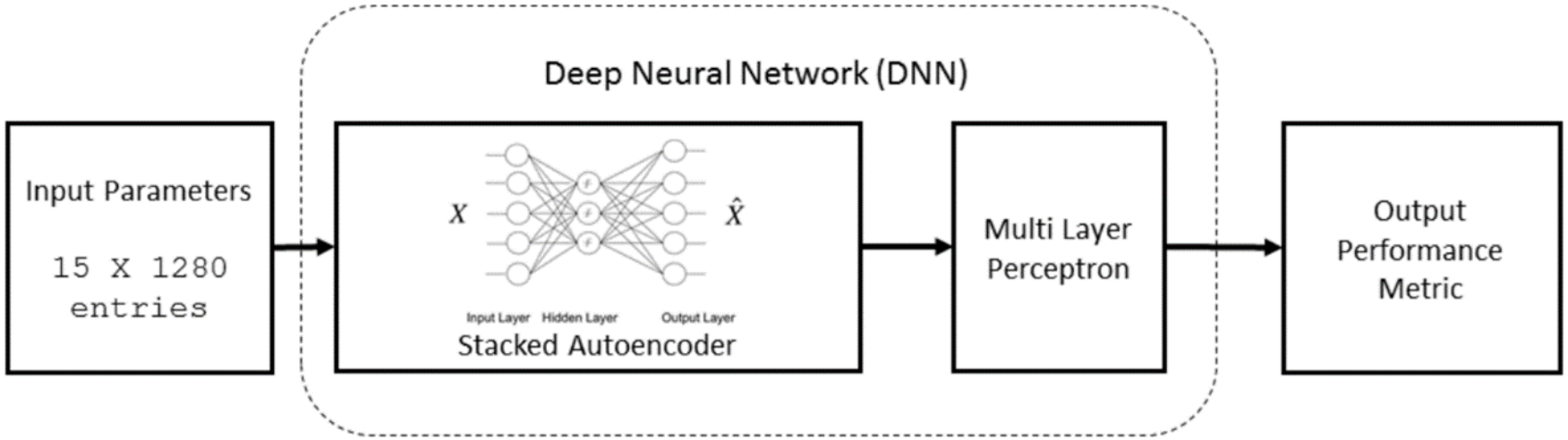




"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Methodology

1. Automated Data Scraping
2. Feature Selection
3. Tag Grouping
4. Feature Learning with Stacked Auto-Encoder
5. Classification by Multilayer Perceptron



"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Automated Data Scraping

- Data from the **GQ India Instagram** account was extracted using the **API** provided by **Instagram**.
- **32 requests per invocation** in a JavaScript Object Notation (JSON)
- **65** features collected for a total of **1280 entries** or **posts**.
  - **1280 X 65: 83200 data-points**
- **Quantization** of data:
  - number of likes in the post were **granulized** to groups of **25**
    - Eg. likes between **0-25** were labeled as **Class 1**, and so on.

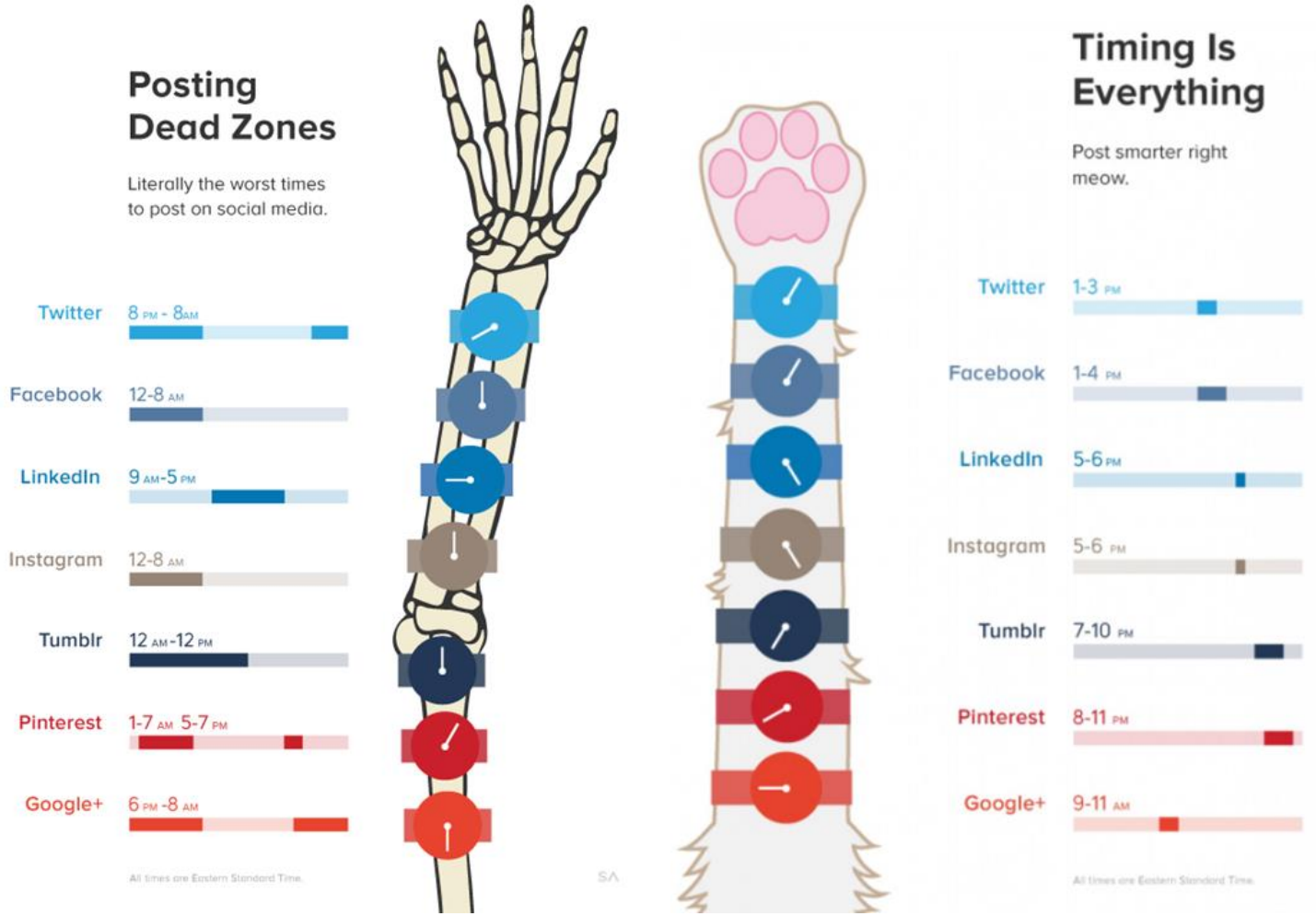
"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Feature Selection

Filter Applied
Creation Time
Week Of The Year
Day of The Week
Hour of the Day
Image (JPG)
Caption
Length of Caption
Number of Tags
Tag List

"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

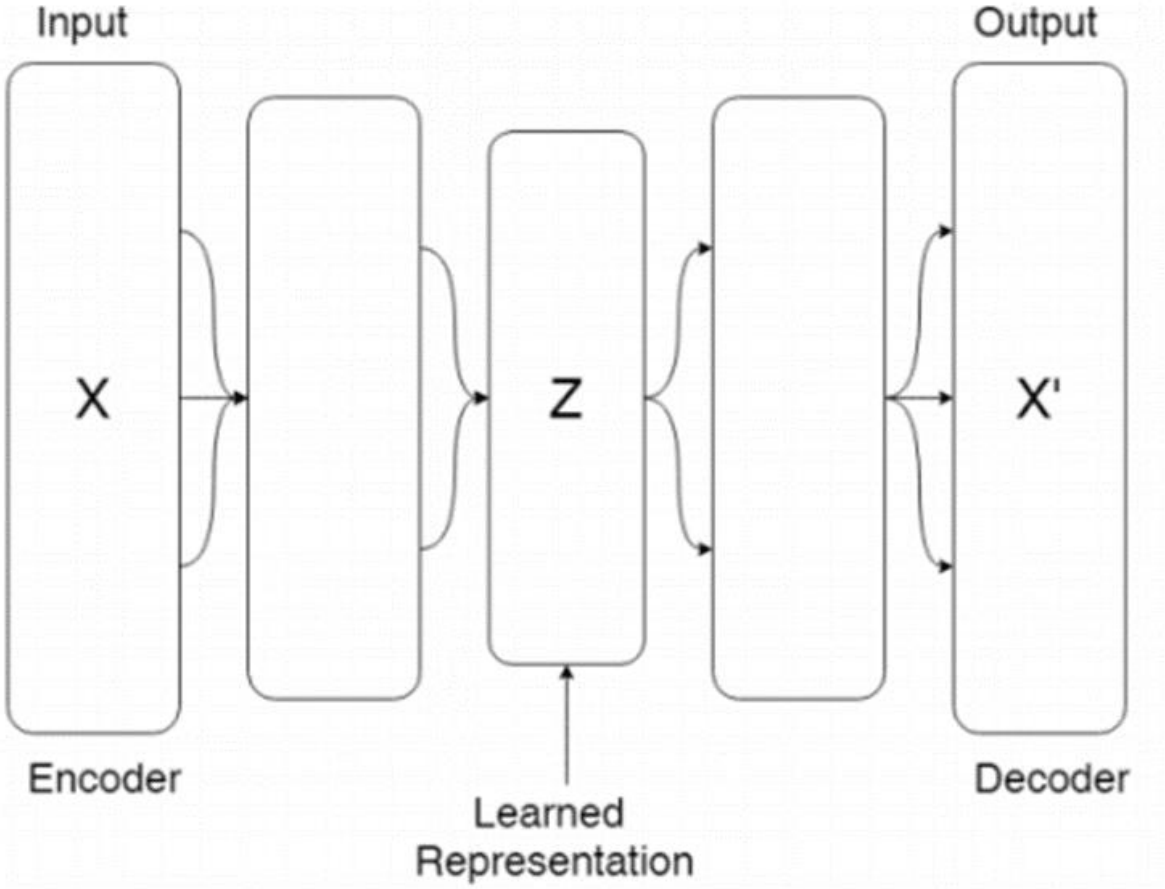
Why is time of the post needed?



"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

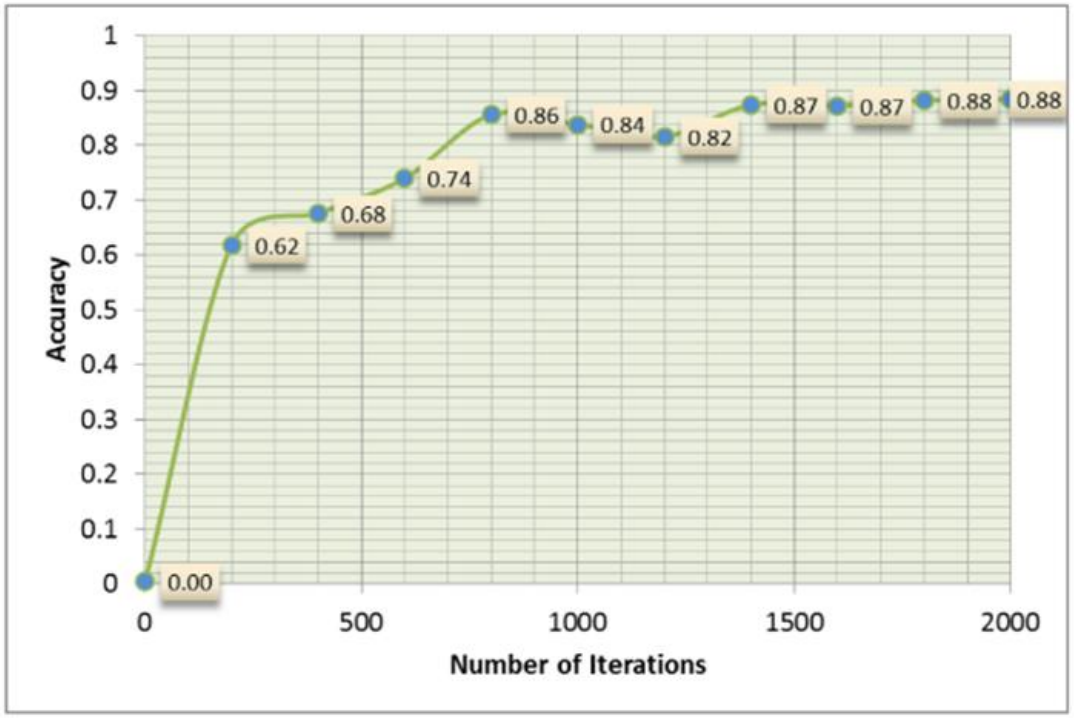
Auto Encoder

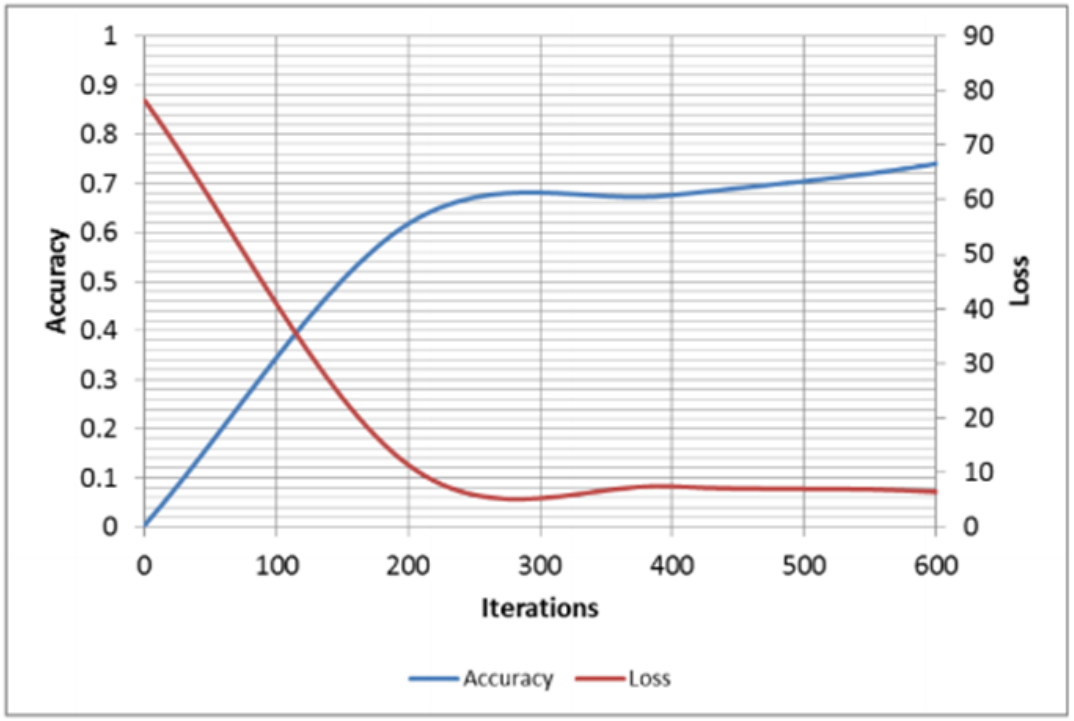
- 4 Layer **stacked autoencoder** is used to obtain an **optimal representation** of the data.
- This is extracted from the 'Z' layer as "**features**".
- Extracted features are classified with a **Multi-layer perceptron**.



"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Results





"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.

Conclusion

Network is able to deliver accuracy of classification higher than 88%.

With a granularity of 25 likes per class, this performance is acceptable for **commercial applications** such as prediction of popularity of a sponsored post, hence price fixation.

In the future, this system can be improved using **computer vision** and **CNN** based techniques to **enrich** the **input features**.

"**Predicting the popularity of Instagram posts for a lifestyle magazine using deep learning**." *In Communication Systems, Computing and IT Applications (CSCITA), 2017 2nd International Conference on*, pp. 174-177. **IEEE**, 2017.