

# Machine Learning Lab task List

1. Hand written digit recognition: Both Bangla and English [[MNIST database](#)] [our dataset]
2. Handwritten character recognition [<http://www.banglalekha.org/dataset/>]
3. **Diabetic Retinopathy**  
A Kaggle dataset consisting of retina images of 17,500 patients (for a total of about 35,000 images) has recently been released ([here](#)). There is a label associated with how bad the damage from diabetes is. The aim of this project is to correctly classify the degree of retinopathy.

Contact: Mike Chrzanowski ([mikechrzanowski@baidu.com](mailto:mikechrzanowski@baidu.com)) [[http://cs229.stanford.edu/projectIdeas\\_2016.html](http://cs229.stanford.edu/projectIdeas_2016.html)]

4. **Malignant tumor classification**  
DDSM dataset is a dataset of mammograms that consists of 1,112 patients. There is an accompanying csv file providing metadata for each photo, including the label (benign or malignant), how severe it is, and the shape of the tumor. The dataset is hosted on Dropbox and can be found [here](#). The aim of this project is to classify tumors as being benign or malignant.

Contact: Darvin Yi ([darvinyi@stanford.edu](mailto:darvinyi@stanford.edu)) [[http://cs229.stanford.edu/projectIdeas\\_2016.html](http://cs229.stanford.edu/projectIdeas_2016.html)]

5. **Sentiment Labelled Sentences Data Set**  
*Download:* [Data Folder](#), [Data Set Description](#)

Abstract: The dataset contains sentences labelled with positive or negative sentiment.

<https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences>

6. Predicting News Sharing on Social Media [[poster](#)] [[report](#)]  
Joseph Johnson, Noam Weinberger [<http://cs229.stanford.edu/projects2016.html>]
7. Enhancing Automated Question Classification. William Mee and Seung-Yeoul Yang. [[pdf](#)]  
[<http://cs229.stanford.edu/projects2010.html>]
8. Sentiment Analysis of Short Texts [<http://www.aclweb.org/anthology/C14-1008>]
9. Automatic Keyword Extraction for Text Summarization in e-Newspapers  
[<http://dl.acm.org/citation.cfm?id=2980442>]
10. Convolutional Neural Networks for Sentence Classification [<https://arxiv.org/abs/1408.5882>]
11. Convolutional Neural Networks for Text Categorization: Shallow Word-level vs. Deep Character-level  
[<https://arxiv.org/abs/1609.00718>]
12. POS Tagger using HMM [hidden Markov model trained using the Baum-Welch algorithm POS tagger]
13. POS Tagger using recurrent neural network [<http://ieeexplore.ieee.org/abstract/document/938396/>]
14. Sentiment Analysis of a person like trump [<http://varianceexplained.org/r/trump-tweets/>]
15. Analysis of BPL Data [similar like <https://www.kaggle.com/vaishaligarg/analysing-ipl-data>]
16. Predicting Movie Revenue from Pre-Release Data[[http://cs229.stanford.edu/proj2015/203\\_report.pdf](http://cs229.stanford.edu/proj2015/203_report.pdf)]
17. A Personalized Company Recommender System for Job Seekers [[poster](#)] [[report](#)]  
Ruixi Lin, Yue Kang, Yixin Cai[[http://cs229.stanford.edu/proj2015/221\\_report.pdf](http://cs229.stanford.edu/proj2015/221_report.pdf)]
18. Predicting A Student's Performance [[poster](#)] [[report](#)]  
Vani Khosla [[http://cs229.stanford.edu/proj2015/241\\_report.pdf](http://cs229.stanford.edu/proj2015/241_report.pdf)]
19. A practical approach to Sentiment Analysis of bangla tweets  
[<http://ieeexplore.ieee.org/abstract/document/7375207/>]
20. Language Identification from Text Documents [[poster](#)] [[report](#)]  
Priyank Mathur, Arkajyoti Misra, Emrah Budur
21. Machine Learning Methods for News Popularity Prediction [[poster](#)] [[report](#)]  
Xuandong Lei, Hongsheng Fang, Xiaoti Hu [[http://cs229.stanford.edu/proj2015/326\\_report.pdf](http://cs229.stanford.edu/proj2015/326_report.pdf)]
22. Project Q: Enron E-mail Dataset  
The Enron E-mail data set contains about 500,000 e-mails from about 150 users. The data set is available here:  
Enron Data  
Project ideas:  
Can you classify the text of an e-mail message to decide who sent it?

23. Recognizing Emotion from Static Images [[report](#)] [[poster](#)]  
Jason Chen, Theodora Chu, Priyanka Sekhar [<http://cs229.stanford.edu/proj2016spr/report/026.pdf>]
24. Aspect-based Sentiment Analysis on Hotel Reviews [[report](#)] [[poster](#)]  
Yangyang Yu [<http://cs229.stanford.edu/proj2016spr/report/032.pdf>]
25. Plant Leaf Recognition [[poster](#)] [[report](#)]
26. A predictor for movie success [2012331008]  
<http://cs229.stanford.edu/proj2013/EricsonGrodman-APredictorForMovieSuccess.pdf>
27. Feature Reduction for unsupervised learning  
<https://pdfs.semanticscholar.org/a5d3/4c04012b4953111891480b4e0656ac4e40d4.pdf>
28. **Detect Distracted Driver** [[poster](#)] [[report](#)]
29. **Twitter US Airline Recommendation Prediction** [[report](#)] [[poster](#)]
30. **Predicting Life Expectancy of Acute Myeloid Leukemia (AML) Patients Based on Gene Expression of Cancer Cells** [[report](#)] [[poster](#)]
31. **Learn To Rate Fine Food** [[report](#)] [[poster](#)]
32. **Predicting Perfume Rate and Popularity** [[report](#)] [[poster](#)]
33. **Modeling Political Identity** [[report](#)] [[poster](#)]
34. **Hotel Recommendation Based on Hybrid Model** [[report](#)] [[poster](#)]
35. **Classification of Transcription Start Sites in the Human Genome** [[report](#)] [[poster](#)]