New! **[KNIME Challenge - MNIST handwriting recognition dataset](https://sakai.unc.edu/portal/tool/15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8/discussionForum/message/dfViewThread" \o " KNIME Challenge - MNIST handwriting recognition dataset)**

[**Evan Galloway (emg33)**](https://sakai.unc.edu/portal/tool/15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8/discussionForum/message/dfViewThread) **(Oct 7, 2015 9:56 AM)** - Read by: 4[[https://sakai.unc.edu/messageforums-tool/images/trans.gif?sakai.tool.placement.id=15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8](javascript:void(0);)Mark as Read](javascript:void(0);) | [[Reply to This Message](https://sakai.unc.edu/portal/tool/15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8/discussionForum/message/dfViewThread)Reply](https://sakai.unc.edu/portal/tool/15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8/discussionForum/message/dfViewThread) | [Email](mailto:emg33@live.unc.edu?subject=Feedback%20on%20KNIME%20Challenge%20-%20MNIST%20handwriting%20recognition%20dataset) | [Grade](https://sakai.unc.edu/portal/tool/15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8/discussionForum/message/dfMsgGrade?forumId=25148&topicId=4503262&messageId=2157211&dialogDivId=dialogDiv&frameId=dialogFrame&gradesSavedDiv=gradesSavedDiv&userId=b154f29b-f16c-4189-9558-de055717d7fd) | [Edit](https://sakai.unc.edu/portal/tool/15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8/discussionForum/message/dfViewThread) | [Delete Message](https://sakai.unc.edu/portal/tool/15a5d5b9-66b8-47f2-b9c2-e38bd836b1e8/discussionForum/message/dfViewThread)

I first spent some time looking around at the various white papers on the KNIME website. For instance, these two were pretty interesting:

https://www.knime.org/files/knime\_web\_knowledge\_extraction.pdf

<https://www.knime.org/knime-applications/outlier-detection-in-medical-claims>

I also came across an interesting source for geospatial data for North Carolina: http://www.nconemap.com/However, I decided that trying to do a geospatial project for this might be too time consuming.

I then wandered over to the kaggle.com site, which as a number of analytics competitions. Some are for actual companies, while others are just to learn. I downloaded the dataset for the Digit Recognizer challenge - <https://www.kaggle.com/c/digit-recognizer>

This is a dataset representing thousands of handwritten numbers. Each pixel is encoded with a value from 0 to 255 representing the gradient from white to black. For example, pretty much the first column of all the numbers is 0, because this represents the edge of the drawing area.

I developed a couple of workflows for this, as you can see in the video, one with a tree classification and one with a support vector machine. I really enjoyed browsing the various nodes and learning what they could do. I also like the visual nature of the workflow design.

Anyhow, once I ran the workflows, I got the results you see in the video. I was only able to get down to about a 15% error rate, which is not great. Here is a list of the various algorithms and the lowest error rates on this dataset:

<http://yann.lecun.com/exdb/mnist/>

If I had more time, it's be interesting to try to get this error lower. In general, I plan to participate in some of the kaggle competitions after the semester is over.

|  |
| --- |
| https://sakai.unc.edu/library/image/sakai/shockwave.gif[KnimeChallenge2.swf](https://sakai.unc.edu/access/content/attachment/c3c781a1-34c9-4ed5-bcd1-f08684222839/Forums/58104365-de76-4128-a71a-889ccfcfc3ce/KnimeChallenge2.swf) |
| https://sakai.unc.edu/library/image/sakai/shockwave.gif[KnimeChallenge.swf](https://sakai.unc.edu/access/content/attachment/c3c781a1-34c9-4ed5-bcd1-f08684222839/Forums/ca669f56-d7de-40ed-8309-0e5bbef37fd3/KnimeChallenge.swf) |