Our Problem:

<http://www.drivendata.org/competitions/8/>

Good for Proposal:

<http://info.asprs.org/publications/proceedings/pecora16/Phillips_R.pdf> (explained why this porject)

Suitability of the Spark framework for data classification:

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=12&cad=rja&uact=8&ved=0CCQQFjABOApqFQoTCO7g7L-798gCFUI6PgodGbkNVg&url=http%3A%2F%2Fcomserv.cs.ut.ee%2Fforms%2Fati_report%2Fdownloader.php%3Ffile%3D6BAAC2BE511A728361A225FF6464CF715DB8B9F1&usg=AFQjCNEBRJljd0VHXLt4DuNHQak3z-bKIg&sig2=l6alZQ2hyHFdftTEI6d2VQ>

(Why used spark for classification?)

Algorithm:

<http://ampcamp.berkeley.edu/5/exercises/image-classification-with-pipelines.html>

<http://blogs.quovantis.com/image-classification-using-apache-spark-with-linear-svm/>

<http://grids.ucs.indiana.edu/ptliupages/publications/Large-Scale%20Image%20Classification%20using%20High%20Performance%20Clustering_v8.2.pdf>

This one has all detailed code of all the algorithms:

<http://stanford.edu/~rezab/sparkworkshop/slides/xiangrui.pdf>

Data Science Lab 5:

<https://github.com/cs109/2015lab5/blob/master/Classification.ipynb>