Is Today a Good Day to Ski?

Kyle Moe, John Wilson

Problem Space

Everyone wants a 'good' day when they going skiing, but what classifies a good day? This requires subjective judgments so we use Machine Learning to predict whether a day is a 'good' day to ski by comparing to previous days. Use data: weather and traffic as certain ski days are considered 'good' when there are two feet of snow in the last 48 hours and low traffic (in our case traffic is analogous to attendance). In summary we attempt to classify a new ski day as 'good' base on how closely it resembles days with similar conditions and then the user can decided whether today is a good day to sk.

Approach

We took a two-staged approach to this problem:

- Gaussian Mixture Clustering:
  - Define k number of clusters
  - Ascribe labels in plain-English to generated clusters
- Predictive Model:
  - Probabilistic classifications into our clusters.
    - What cluster a new day will likely fall into?
  - Simulate data using known attributes
    - Apply K-Nearest Neighbors to "fill in" unknowns
  - Assign most likely labels

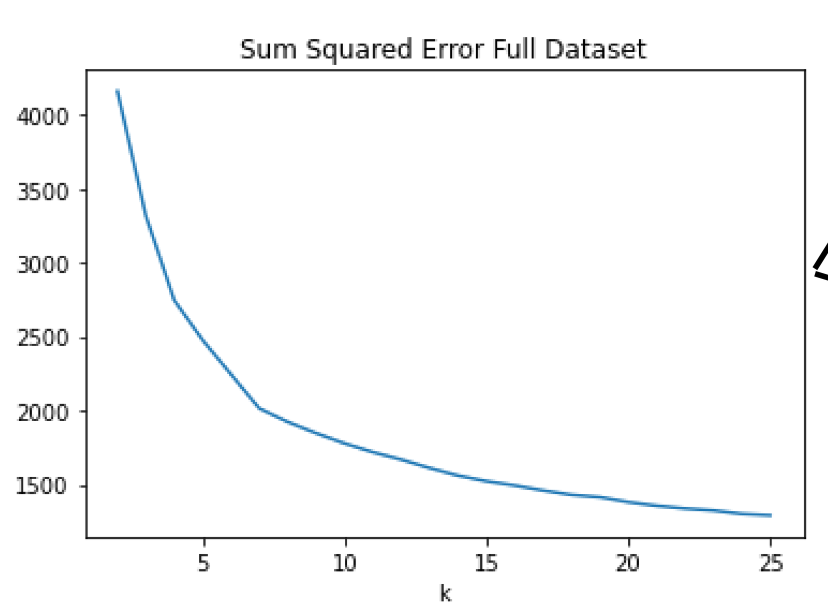
Discussion

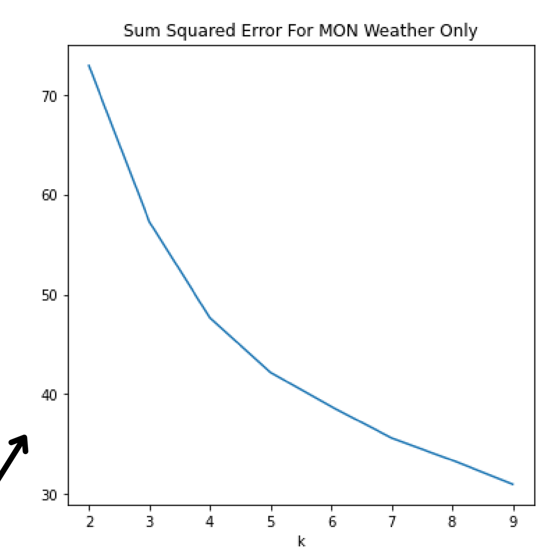
- Clustering provides an effective mechanism to define the quality of a ski day
  - Data from individual days is verbose and changes
  - Much more useful to view labels assigned to clusters
  - The clusters generated are largely unique and it was relatively easy to generate labels for each of the clusters.
- Data challenges
  - Each row corresponded to a single day. This required that we had a sufficient number of days to look at.
    - Measurement methods change between years
  - Could not access raw attendance data. Instead we hsimulated using traffic data which is unreliable at best. This data is useful for relative comparisons between days not as a value itself.
  - Weather data taken from town of Steamboat, not resort
- Predictive model handles simulated data well
  - Can select the correct cluster with relatively high accuracy using only previously know attributes
  - Even greater accuracy if we use 2nd and 3rd most likley clusters from predicted probability table
- Future work
  - Predictive model for raw attendance numbers which requires actual attendance data. Additional improvements to other data sources such as weather can be made as well.
  - Altering prediction scheme to work with datasets of actual predicted weather values rather than simulated data
  - Expand to additional ski resorts

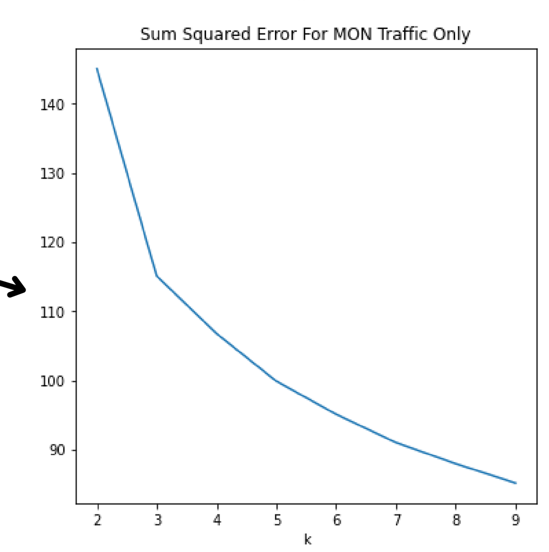
Results

Clustering

- Find cluster count by day -> Approx. 4. (some correlation between traffic and weather)
- Choose to use 30 total clusters
- Apply Gaussian Mixture







Example Labels

Cluster ID	Description	Determination
0	Low recent snow, net negative flux, large 24hr and 48hr flux (Sunday)	High attendance, poor conditions
1 ...	Low recent snow, low daily flux and negative 24/48hr (Thursday)	Low attendance, mediocre conditions
20 ...	High 24/48hr flux, holiday weekend (Monday)	High attendance, variable conditions

Predictive Model

Most likely cluster: 67% accuracy.

Top-two most likely cluster: 88% Accuracy

Top-three most likely clusters: 94% Accuracy

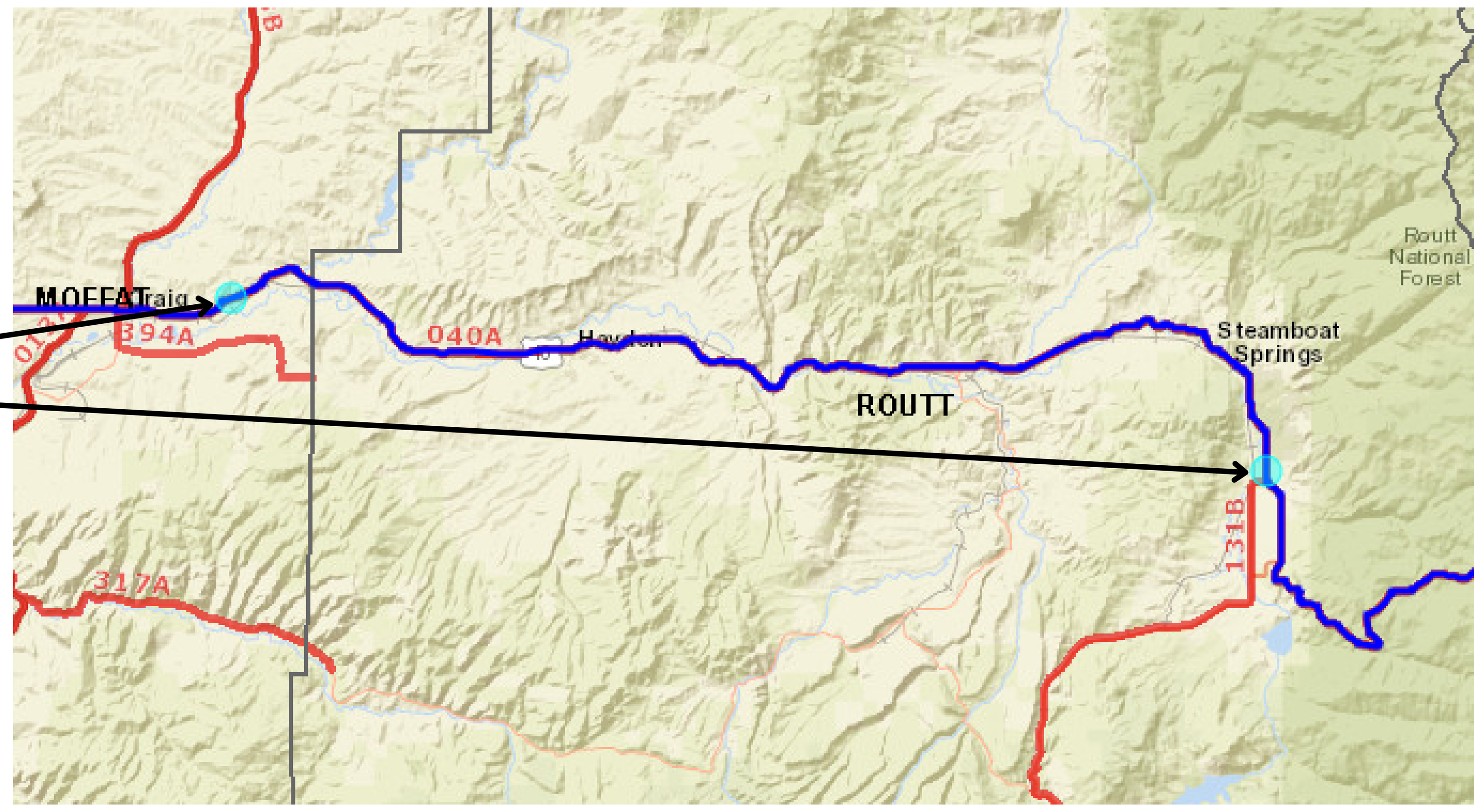
Our approach is effective with simulated data.

Data

Data	Attributes	Derived Attributes	Source
Traffic Data	Net vehicles passing by traffic station in hour period (0-23hr).	Hourly Flux, Daily Flux, and Previous Day's Flux (24hr, 48hr)	Colorado DOT
Weather Data	Snowfall, snowdepth, maximum and minumum temperature	Previous day's snowfall (24hr, 48hr, 7day)	NOAA

Traffic at continuous stations bounding ski resort

Stand-in for attendance data. Difficult to account for locals, alternate routes, and carpooling.



DATE	SNOW	SNWD	TMAX	TMIN	HOUR0	HOUR1	HOUR2	HOUR3	HOUR4	HOUR5	HOUR6	HOUR7	HOUR8	HOUR9	HOUR10	HOUR11	HOUR12	HOUR13	HOUR14	HOUR15
1/22/2003	0	18	31	4	12	18	16	-8	-63	-50	-76	89	104	28	-28	49	23	-4	-23	-25
1/23/2003	1	19	35	26	16	6	8	-5	-59	-60	-102	60	80	42	53	58	0	-20	-43	-21
1/24/2003	1	18	40	28	11	8	1	5	-36	-70	-119	68	86	22	-9	63	-33	-26	36	75

DATE	HOUR16	HOUR17	HOUR18	HOUR19	HOUR20	HOUR21	HOUR22	HOUR23	Flux	PrevFlux2	PrevFlux4	SNOW24	SNOW48	SNOW7	MON	TUE	WED	THU	FRI	SAT	SUN
1/22/2003	68	-39	-32	-33	-63	-41	-4	3	-79	-211	-1288	0	0	2.5	0	0.5	1	0.5	0	0	0
1/23/2003	95	-3	39	14	11	-20	12	5	166	-79	-290	0	0	2.5	0	0	0.5	1	0.5	0	0
1/24/2003	144	104	135	75	97	49	-4	0	682	166	87	1	1	2.5	0	0	0	0.5	1	0.5	0