

Language Classification

Søren Lind Kristiansen

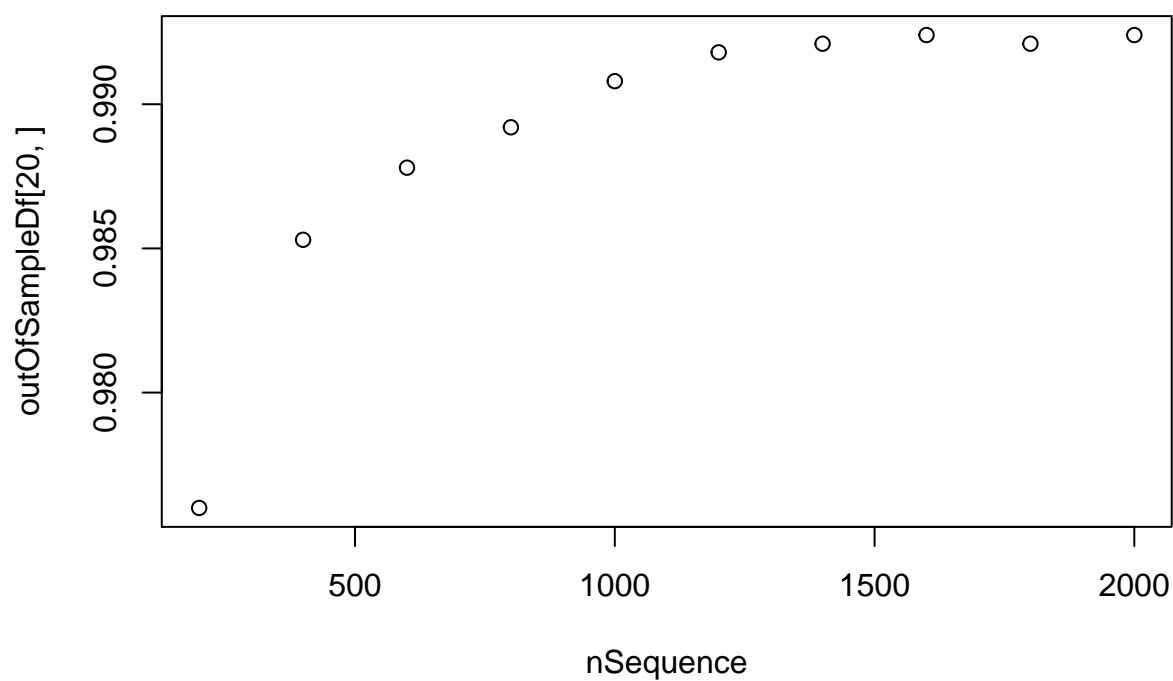
11 December 2015

Feature Selection

```
# {xr dataPreview, cache=TRUE, echo=FALSE} kable(trainingSet[c(25:27,165:167),c(1,6:11,160)],  
align = 'c')
```

Model Training

```
plot(nSequence, outOfSampleDf[20,])
```



Where Is The Beef?

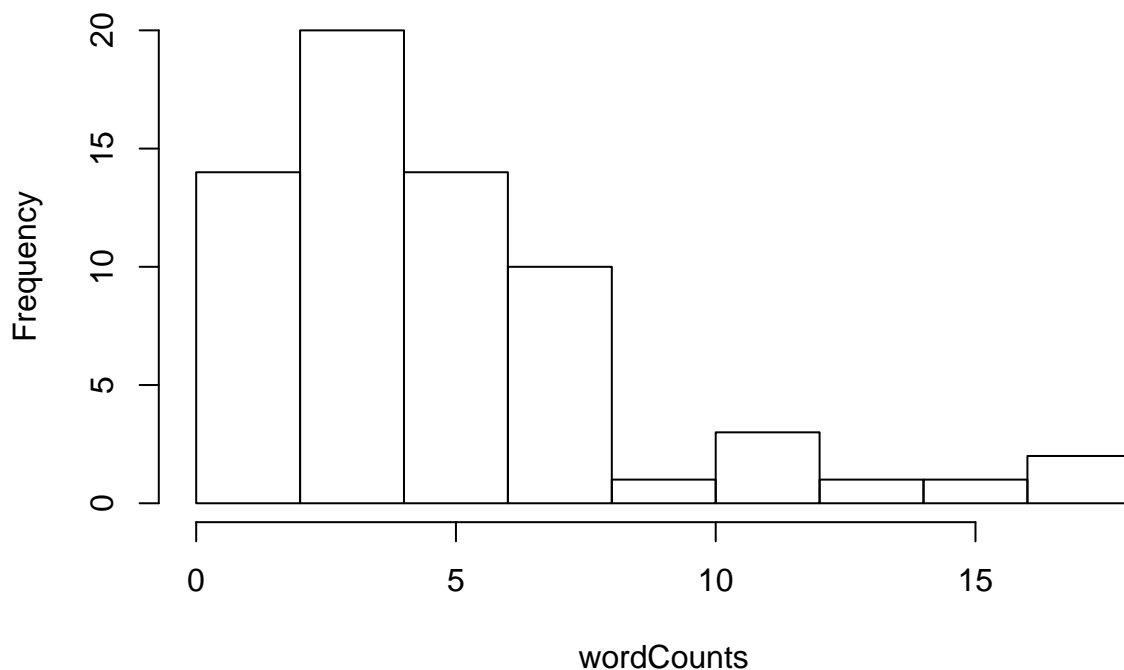
	Prediction	Actual	Text
5108	English	Norwegian	Where is the beef
12355	English	Finnish	Where is the beef

	Prediction	Actual	Text
354	Finnish	French	La renationalisation

	Prediction	Actual	Text
584	Spanish	Portuguese	Estou a referir me especialmente aos anteriores escândalos ligados ao programa Echo
611	Italian	Spanish	No lo creo
613	German	Finnish	Puhemies
688	Spanish	French	Merci Madame Terrón I Cusí
737	Spanish	English	No car manufacturers have been making colossal profits for years
881	Danish	Italian	A ragione
1253	Spanish	Italian	E se sì quanti
1284	English	French	Malheureusement pas grand chose
1594	Dutch	German	Damit ist die Fragestunde an die Kommission beendet
1608	English	Spanish	Theato decisivo
1685	Italian	English	Romano Prodi promised a revolution
1763	English	Italian	E lo dico anche a lei signor Commissario Please deliver the beef even if it is British
1830	Danish	Norwegian	Sedan har vi regeringskonferensen
2007	Portuguese	Spanish	Deseo poder añadir que queremos prestarla
2061	Spanish	English	No
2106	Spanish	Portuguese	Este dinheiro está agora disponível este ano
2157	Danish	Spanish	Informe Rapkay
2323	Spanish	Italian	La seduta sospesa alle riprende alle
2334	Danish	English	Berend Report
2355	Danish	Norwegian	Parlamentet justerade protokollet
2531	English	French	Where is the beef
2687	Finnish	German	Und nun Erika
3024	Danish	Norwegian	Det finner jag inte godtagbart
3491	Dutch	German	Vielen Dank Herr Poettering
3518	Danish	Norwegian	De fick ingen majoritet
3542	Finnish	Norwegian	Detta saknas
3557	Danish	English	COM COS
3638	Dutch	Norwegian	Tack herr Poettering
3765	Danish	Norwegian	Men vi behöver mer öppenhet i konkurrenspolitiken
3926	Danish	Norwegian	De tre s k
3959	Portuguese	Spanish	Cada Presidente o cada Presidenta tiene tantos votos como miembros tiene su Grupo
4027	German	Dutch	Deze eisen gelden dus in alle wateren
4156	German	English	Such is life
4430	Spanish	Portuguese	É fundamental que consigamos obter a fórmula certa
4575	Danish	Norwegian	Applåder Tack fru Roth Behrendt
4964	Spanish	Portuguese	Nós socialistas recusamos esta ingenuidade liberal
5057	Spanish	Portuguese	Aprovo este pedido
5442	English	Spanish	De acuerdo señor Howitt
5734	Danish	German	Sehr geehrte Kollegen

	Prediction	Actual	Text
5806	Dutch	French	Je tiens à féliciter Lousewies van der Laan
6319	English	Spanish	Theato
6520	Danish	Norwegian	Jag skall ge ordet till Provan
6965	Danish	Norwegian	Den var mycket liten
7010	Portuguese	Spanish	Cualquier demócrata puede ver dos problemas fundamentales
7069	Finnish	German	So viel zur Qualität
7294	Dutch	English	I have been fighting for a European environmental coastguard service for almost ten years here in Parliament
7491	Spanish	Norwegian	miljarder euro
7536	Dutch	German	Rina ist jetzt ins Gerede gekommen
7557	Dutch	Finnish	Avaatte HVK n tien
7614	Danish	English	Do forgive me
7724	Portuguese	Spanish	La Directiva CE responde a esta necesidad
8030	Danish	Dutch	Wat is er verbeterd
8213	Danish	Dutch	Ik kom hier straks op terug
8220	Danish	Norwegian	Resultatet ser vi i dag
8681	Danish	Norwegian	Nej
8687	Italian	French	En Allemand scoreboard sonne un peu comme skateboard
8705	Spanish	Portuguese	Finalmente a ausência de transparência
9208	English	Spanish	Theato
9281	Danish	Norwegian	Liksom Aparicio Sánchez tog ni upp avsaknaden av reformer på fiskeområdet
9282	Danish	Norwegian	Naturligtvis behöver vi fartyg med dubbelt skrov
9479	Spanish	Portuguese	O relatório Von Wogau propõe que o controlo desta legitimidade seja reposto ao nível nacional
9507	Danish	German	Danke Frau Terrón i Cusí
9763	Danish	Norwegian	Vi ser fram emot en givande dialog med er om detta
9847	Italian	Danish	Romano Prodi lovede en revolution
10000	Danish	Finnish	Ei ole

Histogram of wordCounts



	n.200	n.400	n.600	n.800	n.1000	n.1200	n.1400	n.1600	n.1800	n.2000
m.1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.2000	0.9995	0.9995	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.3000	0.9997	0.9997	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.4000	0.9998	0.9998	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.5000	0.9998	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.6000	0.9998	1.0000	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.7000	0.9999	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.8000	0.9999	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.9000	0.9999	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.10000	0.9999	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.11000	0.9999	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.12000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m.13000	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.14000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.15000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.16000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.17000	0.9998	0.9999	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.18000	0.9998	0.9999	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.19000	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.20000	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.21000	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.22000	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.23000	0.9997	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	1.0000	0.9999
m.24000	0.9997	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
m.25000	0.9997	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

	n.200	n.400	n.600	n.800	n.1000	n.1200	n.1400	n.1600	n.1800	n.2000
m.1000	0.9452	0.9653	0.9693	0.9721	0.9744	0.9762	0.9758	0.9789	0.9755	0.9772
m.2000	0.9552	0.9741	0.9768	0.9785	0.9830	0.9813	0.9812	0.9832	0.9815	0.9832
m.3000	0.9592	0.9760	0.9795	0.9821	0.9830	0.9845	0.9859	0.9866	0.9872	0.9869
m.4000	0.9644	0.9776	0.9814	0.9826	0.9860	0.9863	0.9868	0.9878	0.9874	0.9884
m.5000	0.9662	0.9794	0.9815	0.9849	0.9860	0.9870	0.9880	0.9877	0.9888	0.9886
m.6000	0.9675	0.9808	0.9828	0.9852	0.9872	0.9888	0.9885	0.9894	0.9892	0.9903
m.7000	0.9686	0.9812	0.9832	0.9856	0.9873	0.9884	0.9883	0.9897	0.9898	0.9905
m.8000	0.9702	0.9817	0.9840	0.9859	0.9875	0.9887	0.9891	0.9895	0.9900	0.9904
m.9000	0.9693	0.9818	0.9840	0.9871	0.9879	0.9895	0.9903	0.9908	0.9903	0.9901
m.10000	0.9715	0.9818	0.9842	0.9857	0.9889	0.9890	0.9896	0.9900	0.9913	0.9903
m.11000	0.9722	0.9834	0.9857	0.9870	0.9892	0.9896	0.9903	0.9911	0.9910	0.9907
m.12000	0.9725	0.9838	0.9848	0.9874	0.9896	0.9898	0.9905	0.9907	0.9912	0.9911
m.13000	0.9742	0.9838	0.9856	0.9882	0.9901	0.9902	0.9904	0.9920	0.9913	0.9922
m.14000	0.9742	0.9835	0.9862	0.9881	0.9898	0.9907	0.9919	0.9917	0.9919	0.9926
m.15000	0.9750	0.9840	0.9857	0.9884	0.9911	0.9906	0.9914	0.9915	0.9919	0.9928
m.16000	0.9761	0.9852	0.9864	0.9892	0.9900	0.9913	0.9909	0.9923	0.9922	0.9929
m.17000	0.9768	0.9859	0.9865	0.9889	0.9906	0.9913	0.9918	0.9919	0.9928	0.9924
m.18000	0.9755	0.9854	0.9867	0.9891	0.9913	0.9911	0.9922	0.9931	0.9926	0.9923
m.19000	0.9762	0.9858	0.9867	0.9894	0.9907	0.9915	0.9917	0.9920	0.9927	0.9924
m.20000	0.9760	0.9853	0.9878	0.9892	0.9908	0.9918	0.9921	0.9924	0.9921	0.9924
m.21000	0.9774	0.9860	0.9881	0.9896	0.9914	0.9920	0.9923	0.9919	0.9922	0.9929
m.22000	0.9767	0.9861	0.9876	0.9896	0.9908	0.9923	0.9919	0.9922	0.9927	0.9931
m.23000	0.9776	0.9866	0.9877	0.9902	0.9914	0.9917	0.9929	0.9920	0.9933	0.9931
m.24000	0.9777	0.9868	0.9884	0.9902	0.9926	0.9919	0.9924	0.9930	0.9934	0.9933
m.25000	0.9771	0.9870	0.9883	0.9899	0.9914	0.9923	0.9930	0.9927	0.9934	0.9936

