

Machine Learning Programming Contest

uitmine

March 1 - 15 , 2013

Do you have it in you to build a system that can categorize tweets?

The TwitMiner 2013 Challenge A Machine Learning Programming Contest

Harikrishna Narasimhan and Saneem Ahmed

TwitMiner Team: Saneem Ahmed Anirban Laha Adway Mitra Harikrishna Narasimhan **Kevin Patel**







- Part of Open Days 2013 @ CSA, IISc
- A machine learning programming contest.

- Part of Open Days 2013 @ CSA, IISc
- A machine learning programming contest.
- Spread awareness about ML among undergrads

- Part of Open Days 2013 @ CSA, IISc
- A machine learning programming contest.
- Spread awareness about ML among undergrads

Challenge

Build a system that can categorize Twitter data.

- Part of Open Days 2013 @ CSA, IISc
- A machine learning programming contest.
- Spread awareness about ML among undergrads

Challenge

Build a system that can categorize Twitter data.

In association with **amazon.com**. Winners to receive **internship** offers!

- Part of Open Days 2013 @ CSA, IISc
- A machine learning programming contest.
- Spread awareness about ML among undergrads

Challenge

Build a system that can categorize Twitter data.

In association with **amazon.com**. Winners to receive **internship** offers!

But first

Why Machine Learning?



Microsoft[®] facebook.





Q

Search

About 28,000,000 results (0.30 seconds)

Web	Scholarly articles for machine learning
Images	in search, optimization, and machine learning - Goldberg - Cited by 50557 Q-learning - Watkins - Cited by 6525
Maps	An introduction to MCMC for machine learning - Andrieu - Cited by 691
Maps	Machine Learning Millingdie, the free energlandie
Videos	Machine learning - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Machine learning
News	Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to
Blogs	➡ List of machine learning Category:Machine learning - Weka - Transduction
Books	
	Machine Learning Coursera
More	www.coursera.org/course/ml
	25 May 2012 – Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has
Bangalore, Karnataka	
Change location	Machine Learning textbook
	www.cs.cmu.edu/~tom/mlbook.html
The web	Machine Learning, Tom Mitchell, McGraw Hill, 1997. cover; Machine Learning is the
Pages from India	study of computer algorithms that improve automatically through experience
5	
More search tools	Machine Learning
	www.springer.com > > Computer Science > Artificial Intelligence
	Machine Learning is an international forum for research on computational approaches to learning. The journal publishes articles reporting substantive results on
	Introduction to Machine Learning

robotics.stanford.edu/~nilsson/mlbook.html From this page you can download a draft of notes I used for a Stanford course on Machine Learning. Although I have tried to eliminate errors, some undoubtedly ...

amazon	Sachin's Amazon.com Today's	Deals Gift Cards Help	
Shop by Department -	Search All -		
Your Amazon.com Your B	rowsing History Recommended Fo	r You Amazon Betterizer Improve Your Recommendations Your Profile Learn More	
Your Amazon.com > F (If you're not Sachin Nagargoje	Recommended for Yo	u and a second	
Just For Today	These recommendat	ions are based on <u>items you own</u> and more.	
Browse Recommended	view: All <u>New Relea</u>	ses Coming Soon	
Recommendations Amazon Instant Video Appliances Appstore for Android Arts, Crafts & Sewing Automotive Baby	1.	Nikon ML-L3 Wireless Remote Control by Nikon (June 17, 2003) Average Customer Review: ★★★★★ ♥ (1,215) In Stock List Price: \$20.50 Price: \$14.95 30 used & new from \$10.00	Add to Cart Add to Wish List
Beauty		it interested 区 ☆☆☆☆☆ Rate this item cause you added Nikon BM-10 LCD Cover for D90 to your Shopping Cart and more (Fix this)	
Books Books on Kindle Camera & Photo Cell Phones & Accessori Clothing & Accessories	2.	Peltor 97010 Ultimate-10 Hearing Protector by Peltor (January 12, 2005) Average Customer Review: ★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★★	
<u>Computers</u> <u>Electronics</u> Grocery & Gourmet Food		List Price: \$ <u>25.99</u> Price: \$20.63	Add to Cart Add to Wish List

45 used & new from \$18.77

I own it Not interested 🛛 🏠 🏠 🏠 Rate this item Recommended because you purchased Peltor 97079 Combat Arms Earplugs and more (Fix this)

з.

Nikon DK-21 Rubber Eyecup Nikon Digital SLR Cameras by Nikon (June 17, 2003) Average Customer Review: ****** 🗹 (28) In Stock Price: \$8.94

7 used & new from \$4.25

Add to Cart Add to Wish List

Music Musical Instruments I own it ONOT interested I ☆☆☆☆☆ Rate this item Recommended because you added Nikon BM-10 LCD Cover for D90 to your Shopping Cart and more (Fix this) Office & School Supplies

Datio Lawn & Cardon

Magazine Subscriptions

Health & Personal Care

Home & Kitchen

Kitchen & Dining

MP3 Downloads

Movies & TV

Jewelry

Home Improvement Industrial & Scientific



Brief Outline

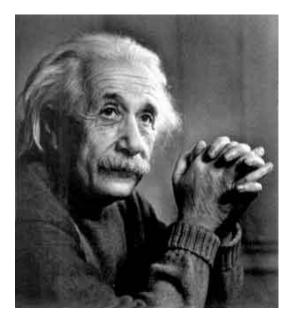
Part 1: Overview of Machine Learning

Brief Outline

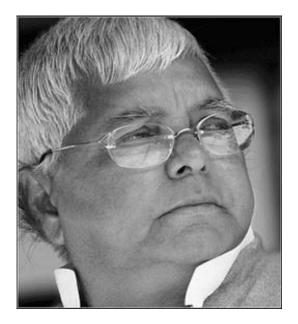
Part 1: Overview of Machine LearningPart 2: The TwitMiner 2013 Challenge

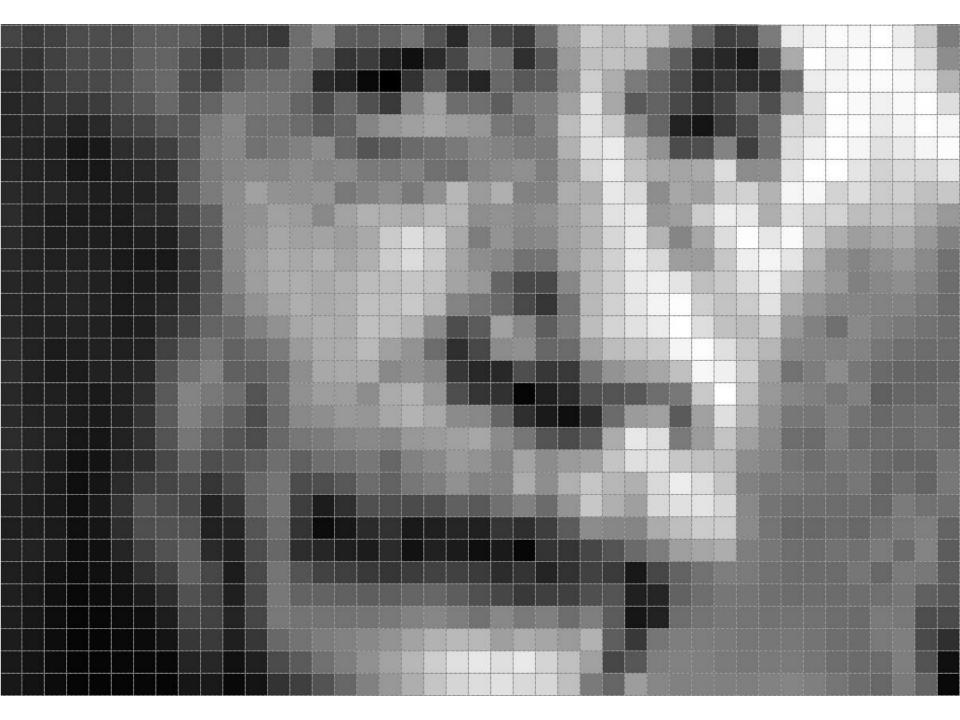
Part 1 What is Machine Learning?

Who among these is Albert Einstein?









																										1			10 0			12					1.1					
63	67	59	77	74	78	92	101	85	71	51	67	74	101	125	99	85	110	112	82	52	97	65	63	120	167	198	198	203	180	151	98	57	66	139	221	254	251	253	250	248	196	124
55	55	59	80	76	84	95	104	78	62	71	80	79	110	98	65	45	31	26	41	66	110	110	50	77	157	198	191	180	134	97	48	39	40	67	167	245	249	252	250	246	217	137
36	44	65	77	61	81	103	98	78	72	93	101	112	117	42	27	8	8	34	87	48	59	101	99	86	144	206	176	127	96	65	43	47	31	53	132	227	250	252	251	250	232	188
38	50	59	53	62	69	87	111	92	93	123	119	114	89	70	116	83	55	123	184	103	91	94	112	128	172	219	159	95	105	75	55	77	93	71	137	233	252	245	248	253	254	230
37	37	31	28	44	64	73	74	87	127	130	130	123	110	98	99	114	144	166	158	150	160	126	116	132	180	233	201	139	99	33	27	43	87	117	194	243	251	246	236	245	247	242
40	43	29	20	28	48	63	64	90	119	124	111	111	129	135	99	92	100	105	112	128	134	119	127	125	178	224	238	175	147	84	68	146	65	113	221	246	251	245	226	229	249	243
40	42	32	24	31	23	28	50	87	133	133	142	133	142	143	127	127	129	117	107	120	127	146	140	120	168	205	241	196	173	164	173	208	150	131	241	249	246	228	214	217	225	235
41	40	35	34	31	32	38	43	86	128	140	167	141	142	141	137	128	136	123	149	177	159	165	139	135	160	186	236	203	153	167	163	204	208	174	250	238	225	220	206	198	191	203
40	39	30	31	27	31	39	44	75	102	139	157	159	153	139	156	168	182	177	187	186	141	144	139	146	163	183	225	228	156	163	208	236	229	172	233	219	207	199	189	184	174	163
39	36	26	32	29	28	30	35	62	93	134	151	167	162	165	165	185	200	204	206	170	124	131	136	136	156	181	216	241	174	137	161	223	252	240	241	197	173	159	159	181	145	120
37	35	30	34	30	25	22	38	46	84	144	156	172	174	166	156	173	202	220	211	163	148	145	125	121	154	184	207	238	216	156	164	191	235	240	229	184	155	132	140	162	138	112
35	36	36	38	35	28	21	46	49	92	137	156	164	178	174	178	179	196	201	185	170	140	92	61	71	135	183	206	228	245	227	196	197	223	241	199	170	149	134	139	138	135	113
35	36	36	34	30	25	25	44	55	91	121	144	170	170	175	181	184	190	194	185	148	109	103	84	55	125	186	213	226	244	243	217	189	202	226	174	146	128	128	136	126	116	111
37	35	32	30	31	23	34	51	68	98	106	126	155	163	164	170	188	196	184	133	72	92	156	145	76	117	190	214	223	234	253	231	185	199	191	144	136	117	130	132	120	106	113
34	35	31	32	40	33	39	70	91	113	101	108	124	151	169	178	184	166	127	131	56	69	133	155	100	103	160	194	198	204	240	250	217	209	177	132	131	128	132	121	115	107	116
40	34	31	27	34	34	44	76	113	132	99	93	112	148	166	178	160	138	156	126	39	48	48	57	66	75	106	146	157	166	198	243	238	210	156	122	126	137	123	101	104	100	107
41	37	31	26	29	34	50	84	129	136	92	87	117	146	158	155	146	169	181	148	75	56	35	23	17	47	82	88	87	112	123	202	239	203	147	130	124	127	122	110	101	102	105
41	40	34	26	28	34	52	87	129	114	88	80	106	133	144	142	157	161	171	180	145	130	108	55	26	26	48	85	162	156	90	162	228	182	139	126	118	121	115	111	103	109	111
41	37	32	22	24	29	46	82	118	98	83	73	101	125	123	141	136	128	156	152	157	165	133	109	77	72	96	186	233	222	123	142	206	165	126	118	114	116	113	104	105	110	114
44	32	28	18	19	34	52	70	94	79	82	97	111	95	103	123	121	101	150	132	156	153	141	154	149	159	163	188	206	228	208	191	188	143	131	116	113	113	113	109	106	102	116
41	32	29	17	18	48	65	66	78	56	70	99	115	79	77	86	106	92	109	130	141	134	133	164	140	178	198	185	192	219	226	231	191	129	128	115	113	113	112	112	113	105	118
32	32	27	15	17	47	78	77	70	40	55	98	121	54	39	46	93	83	64	88	97	79	111	131	92	143	195	170	162	209	225	230	212	139	112	108	109	109	111	111	112	113	111
30	32	26	17	30	42	77	93	75	36	39	84	112	48	8	12	40	53	25	29	37	36	48	40	70	102	145	129	143	166	192	206	196	144	114	109	108	108	110	113	116	125	101
30	32	26	18	21	33	65	86	88	49	34	72	100	52	35	38	27	33	33	28	24	23	28	17	41	48	68	92	98	127	156	172	163	131	118	107	108	109	110	117	122	122	97
28	31	22	10	15	24	44	76	89	68	30	76	117	78	62	66	60	48	60	70	74	50	45	57	45	56	59	48	34	76	98	124	126	120	115	107	107	110	111	116	124	112	101
27	28	19	10	14	18	27	52	76	74	27	68	122	105	104	96	90	92	98	93	96	75	62	66	50	63	87	83	29	25	82	115	115	115	114	111	108	106	111	116	126	92	89
29	24	13	10	11	13	17	29	64	74	27	62	125	116	119	107	123	128	131	133	131	131	133	122	127	126	120	86	32	38	117	120	117	115	116	113	111	109	109	119	125	76	70
24	23	11	7	8	10	10	20	51	67	33	53	110	110	119	133	133	152	167	179	174	175	172	170	161	160	163	103	42	113	128	122	120	120	120	117	115	113	108	116	127	85	47
29	25	13	3	6	10	10	16	30	41	24	34	73	95	115	131	146	167	183	207	224	232	222	223	217	195	160	80	99	133	127	124	122	123	123	123	119	113	110	109	119	92	11
16	23	10	3	5	6	6	12	19	17	14	24	43	63	86	107	125	148	174	188	209	220	223	226	219	184	127	105	137	132	129	127	128	131	131	132	126	117	109	107	115	107	12
								and the second value of th										and the second se	And and a second se	A COLUMN A	1	1	A. C.	A.C. STREET,	And and a second second	And Person in which the		A COLORADOR		and the second se	V. Contraction					A CONTRACTOR OF THE OWNER OWNER OF THE OWNER OWN	A DESCRIPTION OF		The Party number of the Pa	V DOLLAR DE	The second se	The Party number of the Pa

					0	_	_							_							-							£11 0		_				1							1
63	67	59	77	74	78	92	101	85	71	51	67	74	101	125	99	85	110	112	82	52	97	65	63 1	120 1	.67 1	98 19	3 203	180	151	98	57	66	139	221	254	251	253	250	248	196	124
55	55	59	80	76	84	95	104	78	62	71	80	79	110	98	65	45	31	26	41	66	110	110	50	77 1	1.57 1	98 19	L 180	134	97	48	39	40	67	167	245	249	252	250	246	217	137
36	44	65	77	61	81	103	98	78	72	93	101	112	117	42	27	8	8	34	87	48	59	101	99	86 1	.44 2	06 17	5 127	96	65	43	47	31	53	132	227	250	252	251	250	232	188
38	50	59	53	62	69	87	111	92	93	123	119	114	89	70	116	83	55	123	184	103	91	94	112 1	128 1	.72 2	19 15	9 95	105	75	55	77	93	71	137	233	252	245	248	253	254	230
37	37	31	28	44	64	73	74	87	127	130	130	123	110	98	99	114	144	166	158	150	160	126	116 1	132 1	.80 2	33 20	1 139	99	33	27	43	87	117	194	243	251	246	236	245	247	242
40	43	29	20	28	48	63	64	90	119	124	111	111	129	135	99	92	100	105	112	128	134	119	127 1	125 1	.78 2	24 23	8 175	147	84	68	146	65	113	221	246	251	245	226	229	249	243
40	42	32	24	31	23	28	50	87	133	133	142	133	142	143	127	127	129	117	107	120	127	146	140 1	120 1	.68 2	05 24	196	173	164	173	208	150	131	241	249	246	228	214	217	225	235
41	40	35	34	31	32	38	43	86	128	140	167	141	142	141	137	128	136	123	149	177	159	165	139 1	135 1	.60 1	86 23	5 203	153	167	163	204	208	174	250	238	225	220	206	198	191	203
40	39	30	31	27	31	39	44	75	102	139	157	159	153	139	156	168	182	177	187	186	141	144	139 1	146 1	.63 1	83 22	5 228	156	163	208	236	229	172	233	219	207	199	189	184	174	163
39	36	26	32	29	28	30	35	62	93	134	151	167	162	165	165	185	200	204	206	170	124	131	136 1	136 1	156 1	81 21	5 241	174	137	161	223	252	240	241	197	173	159	159	181	145	120
37	35	30	34	30	25	22	38	46	84	144	156	172	174	166	156	173	202	220	211	163	148	145	125 1	121 1	.54 1	84 20	7 238	216	156	164	191	235	240	229	184	155	132	140	162	138	112
35	36	36	38	35	28														_																	149	134	139	138	135	113
35	36	36	34	30	25		\bigcap	7	r	h	V	1		I	Т			، (-	ſ	V	1		IY	- (C		r	א ר		1 1	+	Ω	r	5	128	128	136	126	116	111
37	35	32	30	31	23	'	L	C			I	U		ł		C	: C	1	-1		ľ	U		11					1	J	U	L	C			117	130	132	120	106	113
			1																										_												
34	35	31	32	40	33					_													_					-		-					2	128	132	121	115	107	116
34 40	35 34	31 31	32 27	40 34	33 34					-	Γ	ר	F	26	ב			σ	'n	i	74	Δ	F	- i	n	C.	۲ ر	١٢	n	2)					128 137	132 123	121 101	115 104	107 100	116 107
										-	Γ	C	F	Re	90	C	0	g	n		Z	е	E	- i	n	S	te	j	n	?)					128 137 127	132 123 122	121 101 110	115 104 101	107 100 102	116 107 105
40	34	31	27	34	34	52	87	129	114	88	TC ⁸⁰)	F	R E	20	157)	g	180	145	Z (2	55	26	n	S ⁻	162	21	n	?	228	182	139	126	118	128 137 127 121	132 123 122 115	121 101 110 111	115 104 101 103	107 100 102 109	116 107 105 111
40 41	34 37	31 31	27 26	34 29	34 34	52 46	87	129	114	88	80 73)	133 125	R 144 123	142	157 136) 161 128	9	180 152	145 157	Z 130 165	e	55	26	26 4	S ¹⁸ 81	162	156	90	7	228	182	139	126	118 114	128 137 127 121 116	132 123 122 115 113	121 101 110 111 104	115 104 101 103 105	107 100 102 109 110	116 107 105 111 114
40 41 41	34 37 40	31 31 34	27 26 26	34 29 28	34 34 34	52 46 52	87 82 70	129 118 94	114 98 79	88	80 73 97	106 101 111	133 125 95	R 144 123 103	142 141 123	157 136 121	161 128 101	9 171 156 150	180 152 132	145 157 156	Z 130 165 153	108 133 141	55 109 154 1	26 77 149 1	26 4	18 8:	162	156	90	162	228	182 165 143	139 126 131	126 118 116	118 114 113	128 137 127 121 116 113	132 123 122 115 113 113	121 101 110 111 104 109	115 104 101 103 105 106	107 100 102 109 110 102	116 107 105 111 114 116
40 41 41 41	34 37 40 37	31 31 34 32	27 26 26 22	34 29 28 24	34 34 34 34 29		87 82 70 66	129 118 94 78	114 98 79 56	88	80 73 97 99	106 101 111 115	138 125 95 79	144 123 103 77	142 141 123 86	157 136 121 106	161 128 101 92	171 156 150	180 152 132 130	145 157 156 141	130 165 153 134	108 133 141 133	55 109 154 1 164 1	26 77 149 1 140 1	26 4	18 8:	162	156	90	162	228	182 165 143 129	139 126 131 128	126 118 116 115	118 114 113 113	128 137 127 121 116 113 113	132 123 122 115 113 113 112	121 101 110 111 104 109 112	115 104 101 103 105 106 113	107 100 102 109 110 102 105	116 107 105 111 114 116 118
40 41 41 41 41 44	34 37 40 37 32	31 31 34 32 28	27 26 26 22 18	34 29 28 24 19	34 34 34 29 34	52	87 82 70 66 77		114 98 79 56 40	88 83 82	80 73 97	106 101 111 115 121	138 125 95 79 54	144 123 103 77 39	142 141 123 86 46	157 136 121 106 93	161 128 101 92 83	171 156 150 109 64	180 152 132 130 88	145 157 156 141 97	130 165 153 134 79	108 133 141 133 111	55 109 154 1 164 1	26 77 149 1 140 1 92 1	26 4	18 8:	162 5 233 3 206 5 192	156	90 123 208 226	162 142 191 231	228 206 188 191	182 165 143 129 139	139 126 131 128 112	126 118 116 115 108	118 114 113 113 109	128 137 127 121 116 113 113 109	132 123 122 115 113 113 112 111	121 101 110 111 104 109 112 111	115 104 101 103 105 106 113 112	107 100 102 109 110 102 105 113	116 107 105 111 114 116 118 111
40 41 41 41 41 44 41	34 37 40 37 32 32	31 31 34 32 28 29	27 26 26 22 18 17	34 29 28 24 19 18	34 34 34 29 34 48	52 65		78		88 83 82 70	80 73 97 99 98	106 101 111 115	133 125 95 79 54	144 123 103 77	142 141 123 86	157 136 121 106	161 128 101 92	171 156 150 109 64 25	180 152 132 130 88 29	145 157 156 141 97 37	Z 130 165 153 134 79 36	108 133 141 133 111 48	55 109 154 1 164 1 131	26 77 149 1 140 1 92 1	26 4	18 8:	162 5 233 3 206 5 192	156 222 228 219 209	90 123 208 226	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144	139 126 131 128 112 114	126 118 116 115 108 109	118 114 113 113 109 108	128 137 127 121 116 113 113 109 108	132 123 122 115 113 113 112 111 110	121 101 110 111 104 109 112 111 113	115 104 101 103 105 106 113 112 116	107 100 102 110 102 105 113 125	116 107 105 111 114 116 118 111 101
40 41 41 41 41 44 41 32	34 37 40 37 32 32 32 32	31 31 34 32 28 29 27	27 26 22 18 17 15	34 29 28 24 19 18 17	34 34 34 29 34 48 47	52 65 78	77	78 70	40	88 83 82 70 55	80 73 97 99 98	106 101 111 115 121	133 125 95 79 54	144 123 103 77 39	142 141 123 86 46	157 136 121 106 93	161 128 101 92 83	171 156 150 109 64	180 152 132 130 88	145 157 156 141 97	130 165 153 134 79	108 133 141 133 111 48	55 109 154 1 164 1 131 40	26 77 149 1 140 1 92 1 70 1	26 4 72 9 159 1 178 1 143 1 102 1	18 8:	162 5 233 3 206 5 192	156 222 228 219 209	90 123 208 226 225	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144 131	139 126 131 128 112 114	126 118 116 115 108 109	118 114 113 109 108	128 137 127 121 116 113 109 108 109	132 123 122 115 113 113 112 111 110	121 101 110 111 104 109 112 111 113 117	115 104 101 103 105 106 113 112 116 122	107 100 102 110 102 105 113 125 122	116 107 105 111 114 116 118 111 101 97
40 41 41 41 44 41 32 30	34 37 40 37 32 32 32 32 32	31 31 34 32 28 29 27 26	27 26 22 18 17 15 17	34 29 28 24 19 18 17 30	34 34 34 29 34 48 47 42	52 65 78 77	77 93	78 70 75	40 36	88 83 82 70 55 39	80 73 97 99 98 84	106 101 111 115 121 112	133 125 95 79 54 48	144 123 103 77 39 8	142 141 123 86 46 12	157 136 121 106 93 40	161 128 101 92 83 53	171 156 150 109 64 25	180 152 132 130 88 29	145 157 156 141 97 37	130 165 153 134 79 36	108 133 141 133 111 48	55 109 154 1 164 1 131 40 17	26 77 149 1 140 1 92 1 70 1 41	26 4 72 9 159 1 178 1 143 1 102 1 48 6	18 5: 96 18 63 18 98 18 95 17 45 12	162 5 233 3 206 5 192 0 162 9 143 98	156 222 228 219 209	90 123 208 226 225	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144 131	139 126 131 128 112 114 118 115	126 118 116 115 108 109 107	118 114 113 109 108 108	128 137 127 121 116 113 109 108 109 110	132 123 122 115 113 113 112 111 110 1110	121 101 110 111 104 109 112 111 113 117 116	115 104 101 103 105 106 113 112 116 122 124	107 100 102 109 110 102 105 113 125 122 112	116 107 105 111 114 116 118 111 101 97 101
40 41 41 41 44 41 32 30 30	34 37 40 37 32 32 32 32 32 32	 31 31 34 32 28 29 27 26 26 	27 26 22 18 17 15 17 18	34 29 28 24 19 18 17 30 21	34 34 34 29 34 48 47 42 33	52 65 78 77 65	77 93 86	78 70 75 88	40 36 49	88 83 82 70 55 39 34	80 73 97 99 98 84 72	106 101 111 115 121 112	133 125 95 79 54 48	144 123 103 77 39 8	142 141 123 86 46 12	157 136 121 106 93 40 27	161 128 101 92 83 53	171 156 150 109 64 25 33	180 152 132 130 88 29	145 157 156 141 97 37	130 165 153 134 79 36 23	108 133 141 133 111 48 28	55 109 154 1 164 1 131 40 17 57	26 77 149 1 140 1 92 1 70 1 41	26 4 72 9 159 1 178 1 143 1 102 1 48 6	18 81 96 18 63 18 98 18 95 17 45 12 58 92	162 5 233 3 206 5 192 0 162 9 143 98 34	156 222 228 219 209	90 123 208 226 225	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144 131 120 115	139 126 131 128 112 114 118 115	126 118 116 115 108 109 107 107	118 114 113 109 108 108 107 108	128 137 127 121 116 113 109 108 109 110	132 123 122 115 113 113 112 111 110 110 111	121 101 110 111 104 109 112 111 113 117 116	115 104 101 103 105 106 113 112 116 122 124	107 100 102 109 110 102 105 113 125 122 112 92	116 107 105 111 114 116 118 111 101 97 101 89
40 41 41 41 41 44 41 32 30 30 28	34 37 40 37 32 32 32 32 32 32 32 31	31 31 34 32 28 29 27 26 26 22	27 26 22 18 17 15 17 18 10	34 29 28 24 19 18 17 30 21 15	34 34 34 29 34 48 47 42 33 24	52 65 78 77 65 44	77 93 86 76	78 70 75 88 89	40 36 49 68	88 83 82 70 55 39 34 30	80 73 97 99 98 84 72 76	106 101 111 115 121 112	133 125 95 79 54 48	144 123 103 77 39 8	142 141 123 86 46 12	157 136 121 106 93 40 27	161 128 101 92 83 53	171 156 150 109 64 25 33	180 152 132 130 88 29	145 157 156 141 97 37	130 165 153 134 79 36 23	108 133 141 133 111 48 28	55 109 154 1 164 1 131 40 17 57	26 . 77 . 149 . 140 . 92 . 70 . 41 .	26 4 72 9 159 1 178 1 143 1 102 1 48 6	48 81 96 18 633 18 98 18 95 17 45 12 58 92 59 41	162 5 233 3 206 5 192 0 162 9 143 98 34	156 222 228 219 209 166 127 76	90 123 208 226 225 192 156 98	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144 131 120 115 115	139 126 131 128 112 114 118 115 114 116	126 118 116 115 108 109 107 107 111	118 114 113 109 108 107 108 107	128 137 127 121 116 113 109 108 109 110 106 109	132 123 122 115 113 113 112 111 110 1110 1111 111	121 101 110 111 104 109 112 111 113 117 116 116 119	115 104 101 103 105 106 113 112 116 122 124 126 125	107 100 102 109 110 102 105 113 125 122 112 92 76	116 107 105 111 114 116 118 111 101 97 101 89 70
40 41 41 41 41 41 32 30 30 30 28 27	34 37 40 37 32 32 32 32 32 31 28	 31 31 34 32 28 29 27 26 26 22 19 	27 26 22 18 17 15 17 18 10 10	34 29 28 24 19 18 17 30 21 15 14	34 34 34 29 34 48 47 42 33 24 18	52 65 78 77 65 44 27	77 93 86 76 52	78 70 75 88 89 76	40 36 49 68 74	88 83 82 70 55 39 34 30 27	80 73 97 99 98 84 72 76 68	106 101 111 115 121 112	133 125 95 79 54 48	144 123 103 77 39 8	142 141 123 86 46 12	157 136 121 106 93 40 27	161 128 101 92 83 53	171 156 150 109 64 25 33	180 152 132 130 88 29	145 157 156 141 97 37	130 165 153 134 79 36 23	108 133 141 133 111 48 28	55 109 154 1 164 1 131 40 17 57	26 . 77 . 149 . 140 . 92 . 70 . 41 .	26 4 72 9 159 1 178 1 143 1 102 1 48 6	48 81 96 18 633 18 98 18 95 17 45 12 58 92 59 41	162 233 206 5 192 0 162 9 143 98 34 29	156 222 228 219 209 166 127 76 25	90 123 208 226 225 192 156 98	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144 131 120 115 115 120	139 126 131 128 112 114 118 115 114 116 120	126 118 116 108 109 107 111 113 117	118 114 113 109 108 108 107 108 111 115	128 137 127 121 116 113 109 108 109 100 106 109 113	132 123 122 115 113 113 113 111 110 111 111 109 108	121 101 110 111 104 109 112 111 113 117 116 119 116	115 104 101 103 105 106 113 112 116 122 124 126 125 127	107 100 102 109 110 102 105 113 125 122 112 92 76 85	116 107 105 111 114 116 118 111 101 97 101 89 70 47
40 41 41 41 41 44 41 32 30 30 28 27 29	34 37 40 37 32 32 32 32 32 32 32 31 28 24	31 31 34 32 28 29 27 26 26 26 22 19 13	27 26 22 18 17 15 17 18 10 10	34 29 28 24 19 18 17 30 21 15 14 11	34 34 34 29 34 48 47 42 33 24 18 13	52 65 78 77 65 44 27 17	77 93 86 76 52 29	78 70 75 88 89 76 64	40 36 49 68 74 74	88 83 82 70 55 39 34 30 27 27	80 73 97 99 98 84 72 76 68 68	106 101 111 115 121 112	133 125 95 79 54 48	144 123 103 77 39 8	142 141 123 86 46 12	157 136 121 106 93 40 27	161 128 101 92 83 53	171 156 150 109 64 25 33	180 152 132 130 88 29	145 157 156 141 97 37 24 74 96 131 174	130 165 153 134 79 36 23	108 133 141 133 111 48 28	55 109 154 1 164 1 131 40 17 57	26 . 77 . 149 . 140 . 92 . 70 . 41 .	26 4 72 9 159 1 178 1 143 1 102 1 48 6	48 81 96 18 633 18 98 18 95 17 45 12 58 92 59 41	162 233 206 192 162 162 143 98 34 29 32 32 34	156 222 228 219 209 166 127 76 25	90 123 208 226 225 192 156 98	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144 131 120 115 120 123	139 126 131 128 112 114 118 115 114 116 120 123	126 118 116 115 108 107 107 111 113 117 123	118 114 113 109 108 107 108 111 115 119	128 137 127 121 116 113 109 108 109 110 106 109 113 113	132 123 122 115 113 113 113 112 111 110 111 111 109 108 110	121 101 110 111 104 109 112 111 113 117 116 116 119 116 109	115 104 101 103 105 106 113 112 122 124 126 125 127 119	107 100 102 109 110 102 105 113 125 122 112 92 76 85 92	116 107 105 111 114 116 118 111 101 97 101 89 70 47 11
40 41 41 41 44 41 32 30 30 28 27 29 24	34 37 40 37 32 32 32 32 32 32 31 28 24 23	31 31 34 32 28 29 27 26 26 22 19 13 11	27 26 22 18 17 15 17 18 10 10 10 7	34 29 28 24 19 18 17 30 21 15 14 11 8	34 34 34 29 34 48 47 42 33 24 18 13 10	52 65 78 77 65 44 27 17 10	77 93 86 76 52 29 20	78 70 75 88 89 76 64 51	40 36 49 68 74 74 67	88 83 82 70 55 39 34 30 27 27 33	80 73 97 99 84 72 76 68 62 53	106 101 111 115 121 112	133 125 95 79 54 48	144 123 103 77 39 8	142 141 123 86 46 12	157 136 121 106 93 40 27	161 128 101 92 83 53	171 156 150 109 64 25 33	180 152 132 130 88 29 28 70 93 133 179 207	145 157 156 141 97 24 74 96 131 174 224	130 165 153 134 79 36 23 50 75 131 175 232	108 133 141 133 111 48 45 62 133 172 222	55 109 154 164 131 40 17 66 122 170 122 170 223 2	26 . 77 . 149 . 140 . 92 . 70 . 41 .	26 4 772 5 1,59 1 1,78 1 1,78 1 1,78 1 1,78 1 1,78 1 1,78 1 1,78 1 1,02 1 448 6 556 5 663 6 1,26 1 1,60 1 1,95 1	18 91 96 18 96 18 98 18 98 18 98 18 95 17 95 17 95 17 80 92 95 17 81 12 82 92 81 84 83 84 63 10	162 233 206 192 162 143 98 34 29 32 32 42	156 222 228 219 209 166 127 76 25	90 123 208 226 225 192 156 98	162 142 191 231 230	228 206 188 191	182 165 143 129 139 144 131 120 115 120 123 131	139 126 131 128 112 114 118 115 114 116 120 123 131	126 118 116 108 109 107 111 113 117 123 132	118 114 113 109 108 108 107 108 111 115 119 126	128 137 127 121 116 113 109 108 109 100 106 109 113 113 113	132 123 122 115 113 113 112 111 110 110 111 109 108 110 109	121 101 110 111 104 109 112 111 113 117 116 119 116 109 107	115 104 101 103 105 106 113 112 116 122 124 125 127 119 115	107 100 102 109 110 102 105 113 125 122 122 122 92 76 85 92 107	116 107 105 111 114 114 116 118 111 101 89 70 47 11 12

Machine Learning

"Field of study that gives computers the ability to learn without being explicitly programmed."

- Arthur Samuel

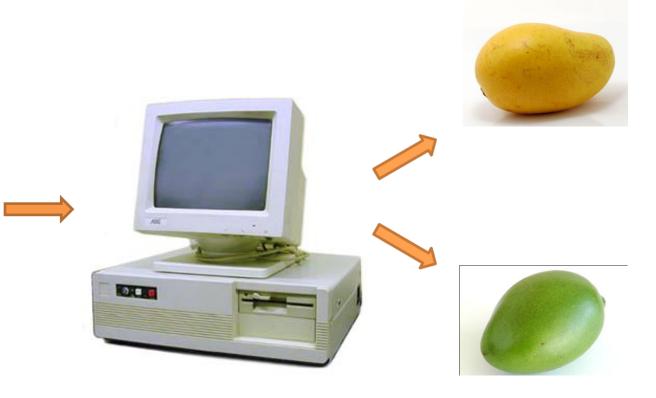




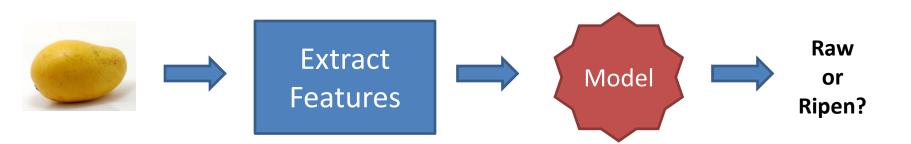




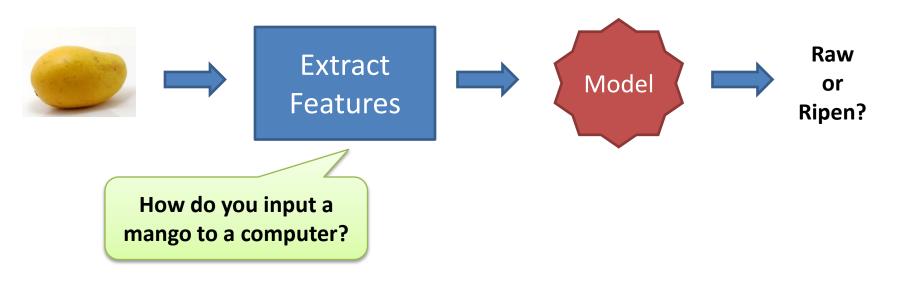




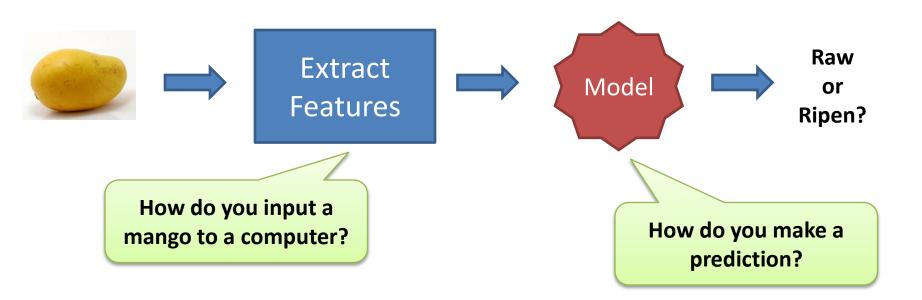
Machine Learning System

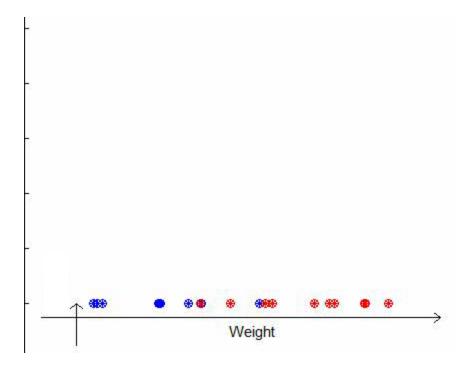


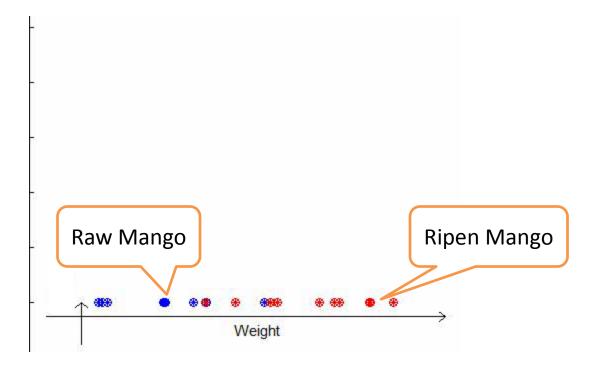
Machine Learning System

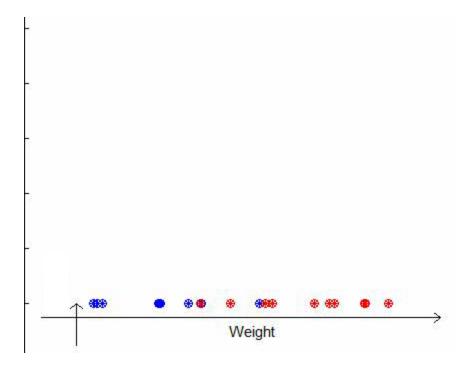


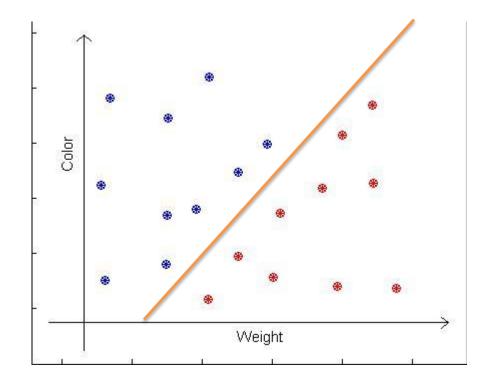
Machine Learning System

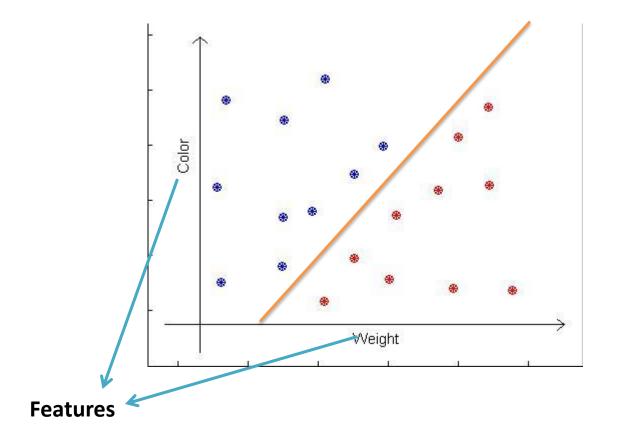


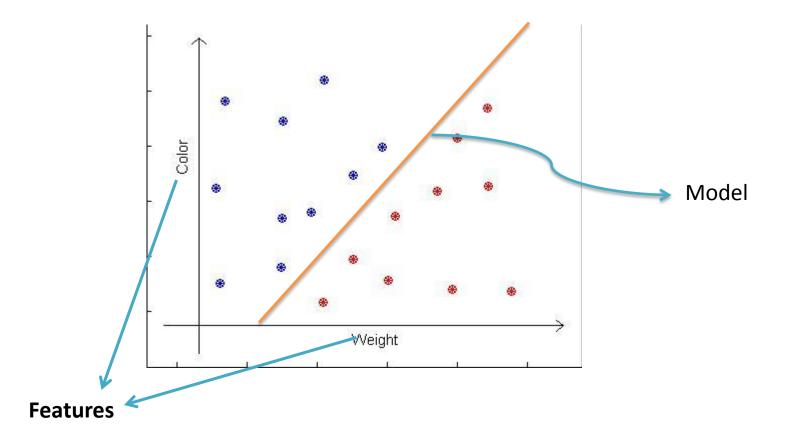












From Allen and Violet Large	Spam x	- E
saavedra saavedra@filo.uba.ar to undisclosed recipients 💌		May 9 🔆 🔺 🔫

A Be careful with this message. Similar messages were used to steal people's personal information. Unless you trust the sender, don't click links or reply with personal information. Learn more

Dear Sir/Madam,

This is my fifth times of writting you this email since last year till date but no response from you. Hope you get this one, as this is a personal email directed to you. My wife and I won a Jackpot Lottery of \$11.3 million in July and have voluntarily decided to donate the sum of \$500,000.00 USD to you as part of our own charity project to improve the lot of 10 lucky individuals all over the world. If you have received this email then you are one of the lucky recipients and all you have to do is get back with us so that we can send your details to the payout bank. Please note that you have to contact my private email for more informations (allenvioletlarge2@ozledim.net)

You can verify this by visiting the web pages below.

http://www.dailymail.co.uk/news/article-1326473/Canadian-couple-Allen-Violet-Large-away-entire-11-2m-lottery-win.html

http://www.cbc.ca/news/canada/nova-scotia/story/2010/11/04/ns-allen-violet-large-lottery-winning.html

Goodluck, Allen and Violet Large Email: allenvioletlarge2@ozledim.net

May 9 📩 🔸 🝷

A Be careful with this message. Similar messages were used to steal people's personal information. Unless you trust the sender, don't click links or reply with personal information. Learn more

Dear Sir/Madam,

This is my fifth times of writting you this email since last year till date but no response from you. Hope you get this one, as this is a personal email directed to you. My wife and I won a Jackpot Lottery of \$11.3 million in July and have voluntarily decided to donate the sum of \$500,000.00 USD to you as part of our own charity project to improve the lot of 10 lucky individuals all over the world. If you have received this email then you are one of the lucky ecipients and all you have to do is get back with us so that we can send your details to the payout bank. Please note that you have to contact my private email for more informations (allenvioletlarge2@ozledim.net)

You can verify this by visiting the web pages below.

http://www.dailymail.co.uk/news/article-1326473/Canadian-couple-Allen-Violet-Large-away-entire-11-2m-lottery-win.html

http://www.cbc.ca/news/canada/nova-scotia/story/2010/11/04/ns-allen-violet-large-lottery-winning.html

Goodluck, Allen and Violet Large Email: allenvioletlarge2@ozledim.net

From	Allen and Violet Large 🕞 Spam x		ē I	2
+	saavedra saavedra@filo.uba.ar to undisclosed recipients	May 9 🕁	•	•
▲	Be careful with this message. Similar messages were used to steal people's personal information. Unless you trust the sender, don't click links or reply with personal information. Learn more			
	-			
	Dear Sir/Madam,			
	This is my fifth times of writting you this email since last year till date but no response from you. Hope you get this one, as this is a personal email directed to you. My wife and I won a Jackpot Lottery of \$11.3 million in July and have voluntarily decided			
	to donate the sum of \$500,000.00 USD to you as part of our own charity project to improve the lot of 10 lucky individuals all over the world. You have received this email then you are one of the			
•	lucky ecipients and all you have to do is get by the us so that we can send your details to the payout bank. Please note that you have to contact my private email re informations (allenvioletlarge2@ozledim.net)			
	You can verify this by visiting the web pages below			
	http://www.dailymail.co.uk/news/article-1326473/Can Look for key words			
	http://www.cbc.ca/news/canada/nova-scotia/story/2010/11/04/ns-allen-violet-large-lottery-winning.html			
	Coodluck			

Goodluck, Allen and Violet Large Email: allenvioletlarge2@ozledim.net

Spam Filter

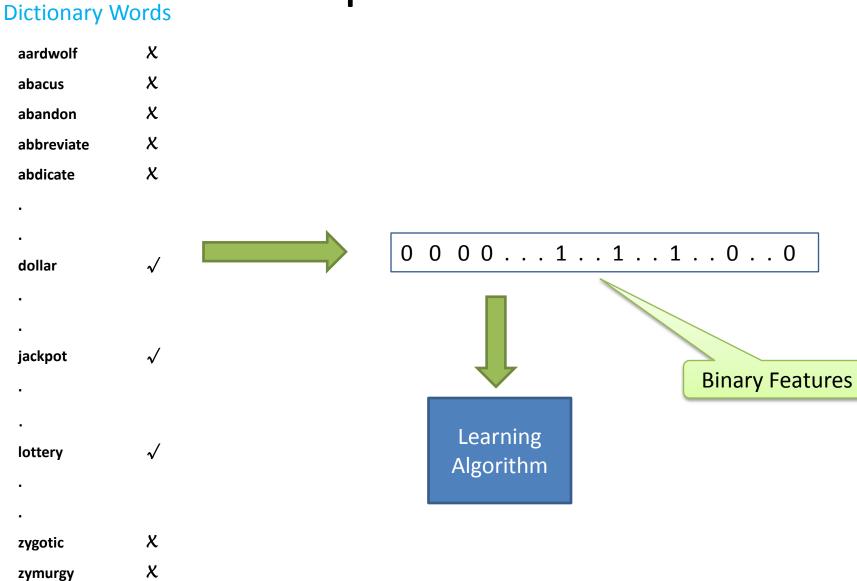
Dictionary Words

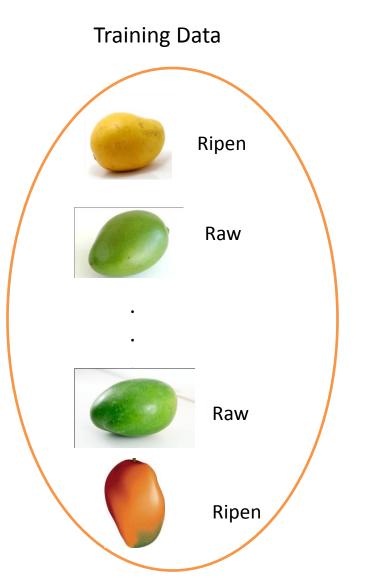
aardwolf	x
abacus	x
abandon	x
abbreviate	x
abdicate	X
dollar	\checkmark
jackpot	\checkmark
lottery	\checkmark
zygotic	X
zymurgy	X

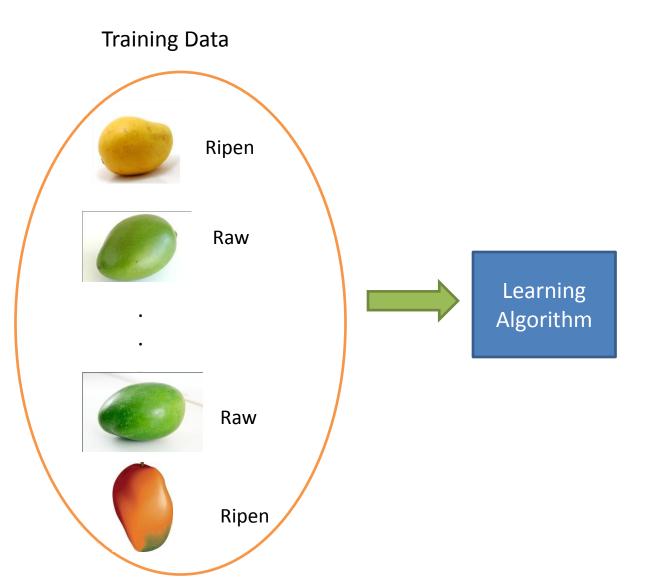
Spam Filter

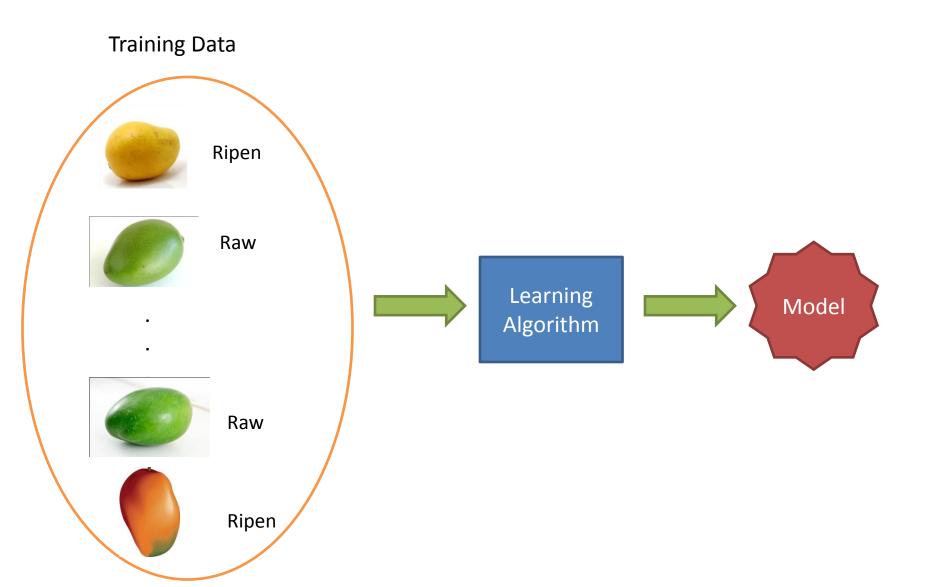
Dictionary Words X aardwolf X abacus X abandon X abbreviate X abdicate • . $0 \ 0 \ 0 \ 0 \ \dots \ 1 \ \dots \ 1 \ \dots \ 1 \ \dots \ 0 \ \dots \ 0$ dollar \checkmark • . jackpot \checkmark **Binary Features** • • \checkmark lottery ٠ ٠ X zygotic X zymurgy

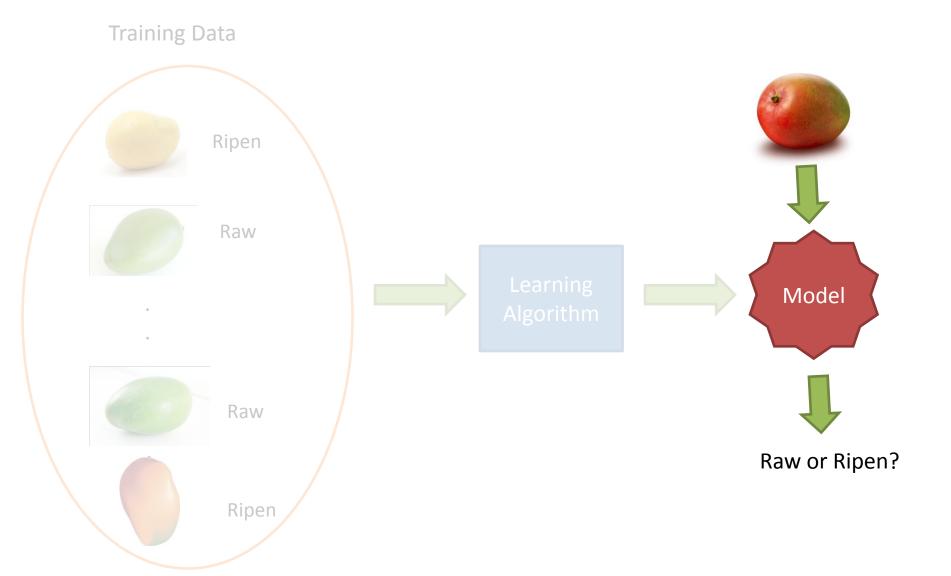
Spam Filter



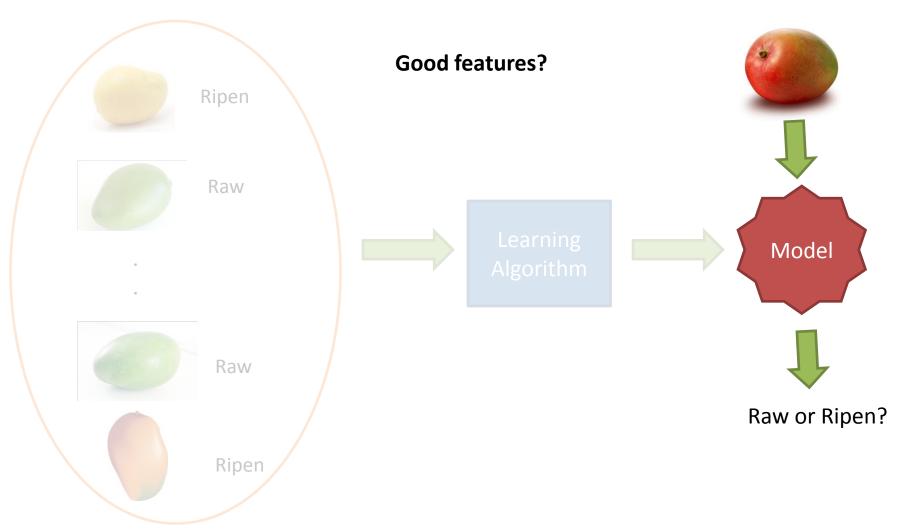




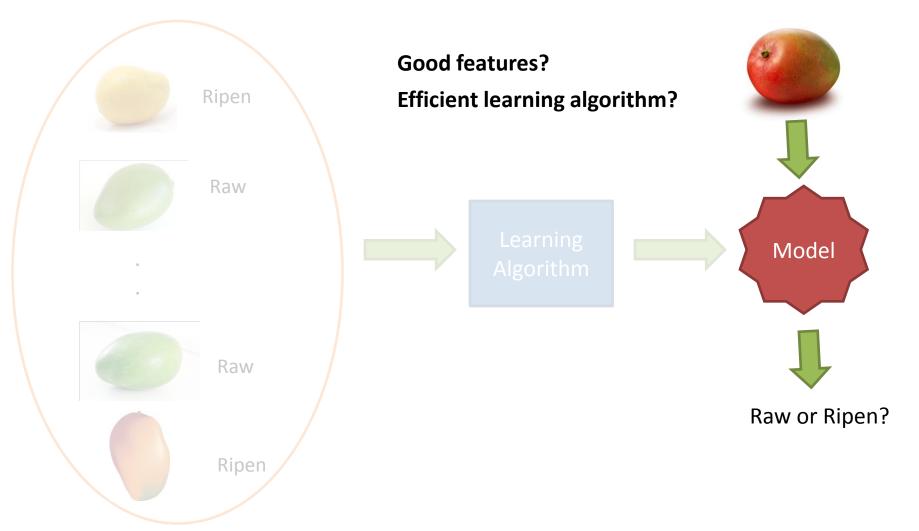


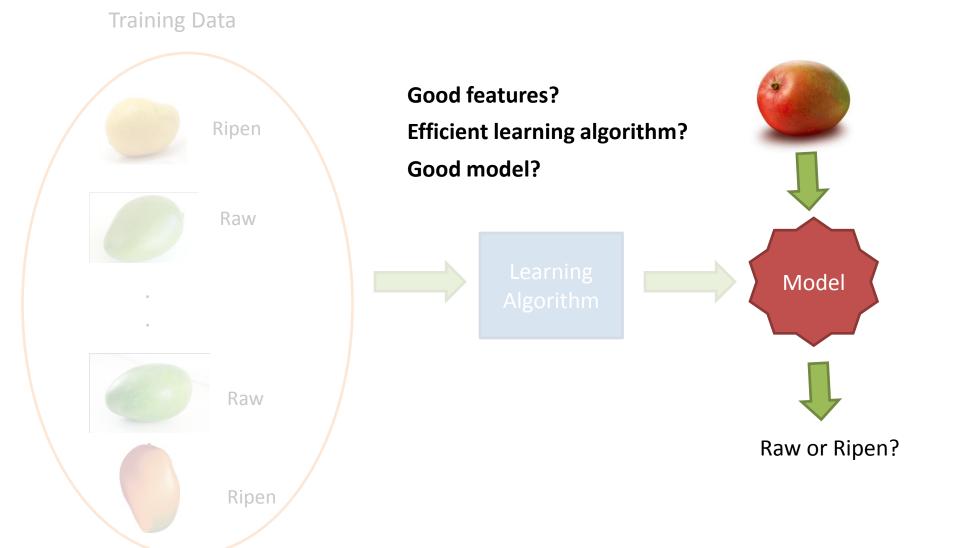


Training Data

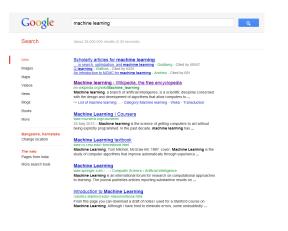


Training Data

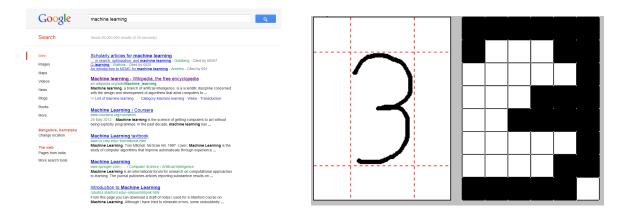




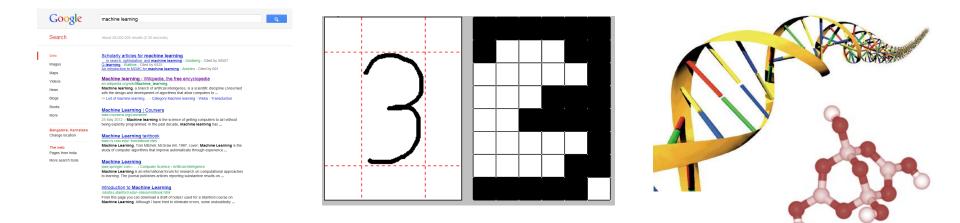
• Information Retrieval



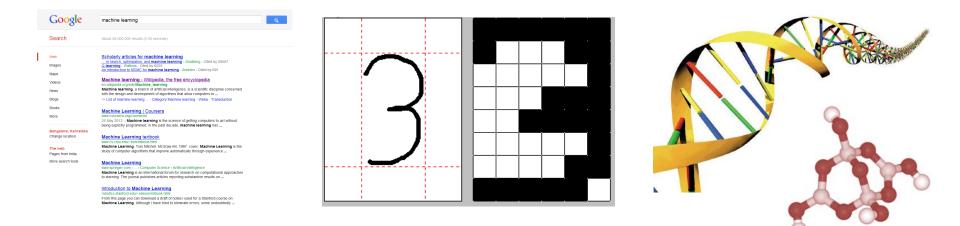
- Information Retrieval
- Computer Vision



- Information Retrieval
- Computer Vision
- Bioinformatics



- Information Retrieval
- Computer Vision
- Bioinformatics
- Natural Language Processing



Part 2 The TwitMiner Challenge 2013

The Challenge

 Build a system that can categorize tweets into sports and politics.

ROONEY	Wayne Rooney @WayneRooney Feb 1 Click here to check out my exclusive interview ahead of tomorrow's game against Real Madrid on my Facebook page - on.fb.me/Xys5d3 Expand	sports
á	Wayne Rooney (programmys on Expand Expand	politics
	Barack Obama @BarackObama Peb 1 "America only moves forward when we do so together—when we accept certain obligations to one another." OFA.BO/NmaATn View media	6

The Challenge

• Build a system that can categorize tweets into sports and politics.

_		
WAYNE	Wayne Rooney @WayneRooney Feb 12 Click here to check out my exclusive interview ahead of tomorrow's game against Real Madrid on my Facebook page - on.fb.me/Xys5d3 Expand	sports
æ	Wayne Rooney (projectione) Feb 11 What channel is grammys on Expand	politics
	Barack Obama @BarackObama Peb 16 "America only moves forward when we do so together—when we accept certain obligations to one another." OFA.BO/NmaATn View media	

Data in a single tweet is limited by number of characters.

• Training data set: <tweet-id, user-id, label>

- Training data set: <tweet-id, user-id, label>
- Validation data set: <tweet-id, user-id >

- Training data set: <tweet-id, user-id, label>
- Validation data set: <tweet-id, user-id >
- Test data set: <tweet-id, user-id >

Used to build a prediction model

- Training data set: <tweet-id, user-id, label>
- Validation data set: <tweet-id, user-id >
- Test data set: <tweet-id, user-id >

Used to build a prediction model

- Training data set: <tweet-id, user-id, label>
- Validation data set: <tweet-id, user-id >
- Test data set: <tweet-id, user-id >

Used to evaluate your model

Used to build a prediction model

- Training data set: <tweet-id, user-id, label>
- Validation data set: <tweet-id, user-id >
- Test data set: <tweet-id, user-id >

Used to evaluate your model

Tweet text not provided!

A separate script provided to download tweet texts.

Used to build a prediction model

- Training data set: <tweet-id, user-id, label>
- Validation data set: <tweet-id, user-id >
- Test data set: <tweet-id, user-id >

Used to evaluate your model

Tweet text not provided!

A separate script provided to download tweet texts.

Data to be uploaded soon.. (before March 3)



Wayne Rooney @WayneRooney Feb 12 Click here to check out my exclusive interview ahead of tomorrow's game against Real Madrid on my Facebook page - on.fb.me/Xys5d3 Expand

1. Extract features from tweet text



Wayne Rooney @WayneRooney Feb 12 Click here to check out my exclusive interview ahead of tomorrow's game against Real Madrid on my Facebook page - on.fb.me/Xys5d3 Expand

- 1. Extract features from tweet text
- 2. Build a model using a learning algorithm



Wayne Rooney @WayneRooney Feb 12 Click here to check out my exclusive interview ahead of tomorrow's game against Real Madrid on my Facebook page - on.fb.me/Xys5d3 Expand

- 1. Extract features from tweet text
- 2. Build a model using a learning algorithm
- 3. Tune the model using validation data



Wayne Rooney @WayneRooney Feb 12 Click here to check out my exclusive interview ahead of tomorrow's game against Real Madrid on my Facebook page - on.fb.me/Xys5d3 Expand

- 1. Extract features from tweet text
- 2. Build a model using a learning algorithm
- 3. Tune the model using validation data
- 4. Submit predicted labels on test data

Submissions

- Submission portal to be online soon..
- Predicted labels for validation data — Helps you tune your model
- Predicted labels for test data

- Used to evaluate your model

Any number of times!

Only once!

Prizes

- First Prize: Internship/PPI offers at amazon.com.
- Second Prize: Prizes worth INR 10000
- Third Prize: Prizes worth INR 6000

• Each team can have a max. of 2 members.

- Each team can have a max. of 2 members.
- Event open from March 1-15.

- Each team can have a max. of 2 members.
- Event open from March 1-15.
- Teams to upload a write up on their approach with properly documented code.

- Each team can have a max. of 2 members.
- Event open from March 1-15.
- Teams to upload a write up on their approach with properly documented code.
- Both test accuracy and submitted write up/code will be used to decide the winner.

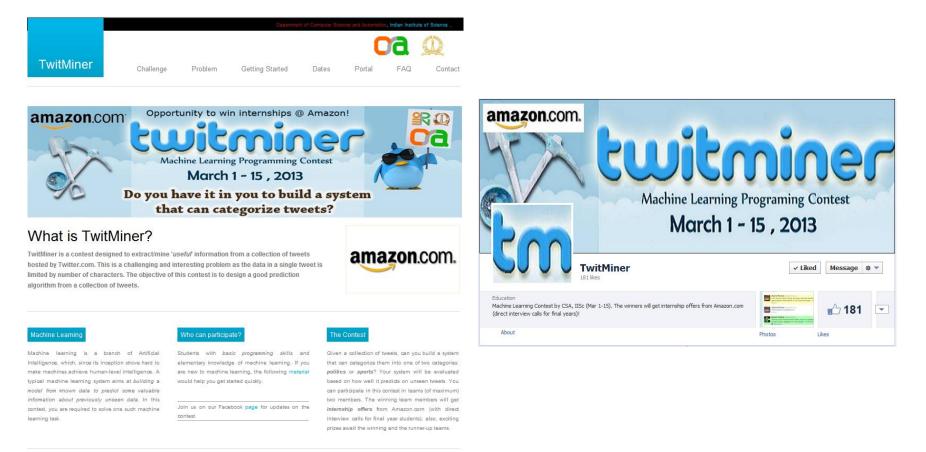
• 1-Mar 2013: Event declared open

- 1-Mar 2013: Event declared open
- 3-Mar 2013: Registration begins

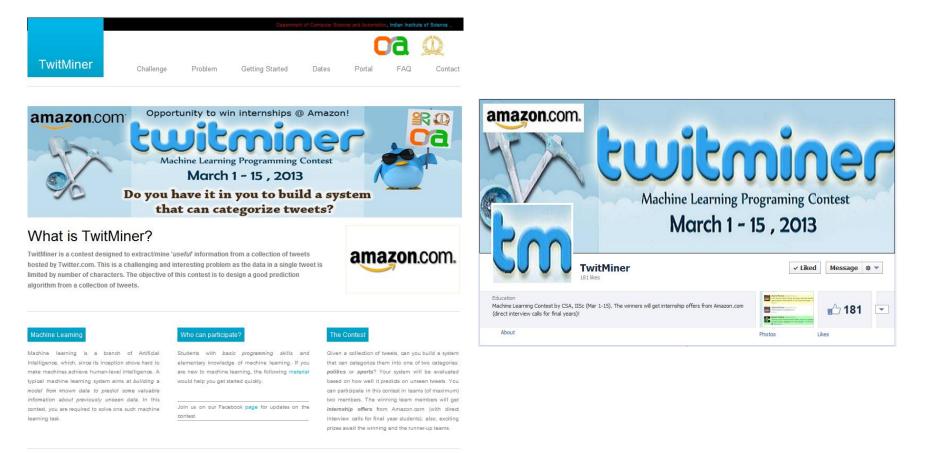
- 1-Mar 2013: Event declared open
- 3-Mar 2013: Registration begins
- 15-Mar 2013:
 - IST 12:00 AM: Test data made available
 - IST 11:59 PM: Event comes to an end

- 1-Mar 2013: Event declared open
- 3-Mar 2013: Registration begins
- 15-Mar 2013:
 - IST 12:00 AM: Test data made available
 - IST 11:59 PM: Event comes to an end
- 31-Mar 2013: Declaration of winners

Visit us at: http://events.csa.iisc.ernet.in/opendays2013/twitminer/



Visit us at: http://events.csa.iisc.ernet.in/opendays2013/twitminer/



Contact Us: twitminer2013@gmail.com Facebook: http://www.facebook.com/twitminer

Registrations open from Mar 3rd...

		Department of Computer Science and Automation, Indian Institute of Science .				
					Ca	Ŵ
TwitMiner	Home	Challenge	Problem	Important Dates	Portal	Contact
Email or login	Ŷ					
Password	Ų					
Remember me Forgot password Register						
Let me in						

Copyright 2013 TwitMiner

Registrations open from Mar 3rd...

		Department of Computer Science and Automation, Indian Institute of Science .				
					Ca	
TwitMiner	Home	Challenge	Problem	Important Dates	Portal	Contact
Email or login Password Remember me Forgot pa Let me in	্ ssword Register					

Copyright 2013 TwitMiner

Questions?