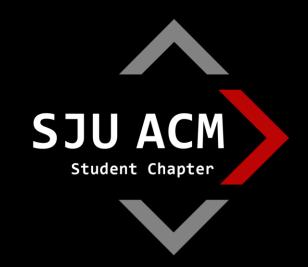
## INTRO TO AI & MACHINE LEARNING





#### WHAT IS MACHINE LEARNING?

- In simplest terms, ML refers to the methods that allow computers to find patterns in data
- By finding patterns in data, ML can be used to form predictions or recommendations





#### IMPORTANCE OF DATA

- Companies look to capture as many data points on their users as possible – this data is extremely valuable
- Companies like Google, Facebook, and Apple have so much of our personal data that they understand us better than we do





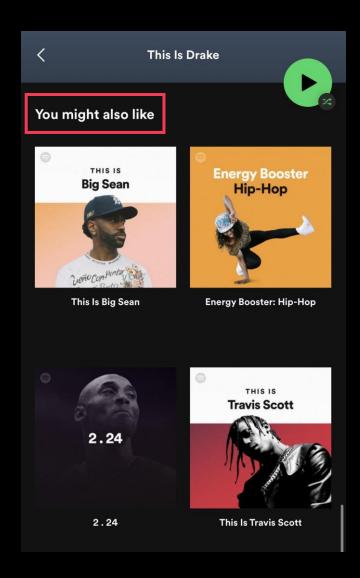
#### **MACHINE LEARNING IS EVERYWHERE!**

- Most common applications of ML:
  - Instagram Home & Explore Page
  - TikTok For You Page
  - Snapchat Filters
  - Amazon Product Suggestions
  - LinkedIn Jobs & Connections
  - Spotify Music Recommendations
  - YouTube Video Recommendations





#### **SPOTIFY RECOMMENDATIONS**









#### **SPOTIFY RECOMMENDATIONS**

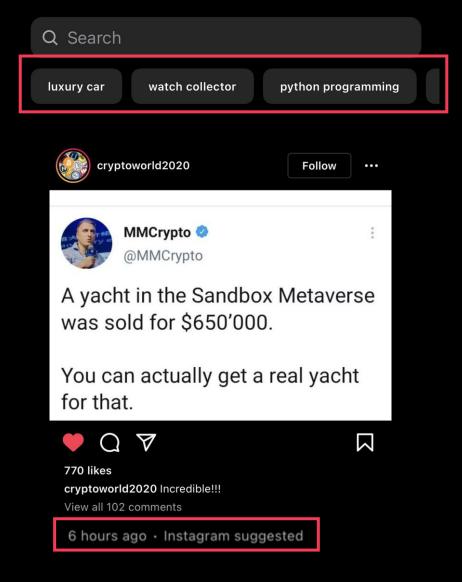
#### How do we use ML?

It all starts with data. At the most fundamental level, all sorts of user info — playlists, listening history, interactions with Spotify's UI, etc. — are fed into our ML models, while keeping trust and responsibility top of mind. Every day, nearly half a trillion events are processed, and the more info our models gather, the smarter they become about making associations between different artists, songs, podcasts, and playlists.



#### **INSTAGRAM EXPLORE PAGE**

Instagram
knows exactly
what my
interests are!







#### **INSTAGRAM EXPLORE PAGE**

# Instagram explains how it uses AI to choose content for your Explore tab

A peek behind the algorithmic scenes

By James Vincent | Nov 25, 2019, 10:00am EST

To make its recommendations, the Explore system begins by looking at "seed accounts," which are accounts that users have interacted with in the past by liking or saving their content. It identifies accounts similar to these, and from them, it selects 500 pieces of content. These candidates are filtered to remove spam, misinformation, and "likely policy-violating content," and the remaining posts are ranked based on how likely a user is to interact with each one. Finally, the top 25 posts are sent to the first page of the user's Explore tab.





### How TikTok's algorithm works

TikTok's algorithm fuels the app's recommendation system that determines which videos will appear on the For You page. Each person's feed is unique: TikTok says factors taken into account include videos you have liked or shared and comments you've made; video data such as hashtags and captions; and your device and account settings.

TikTok uses machine learning based techniques to understand what you view and why. The goal is to keep you on the platform for as long as possible and collect more data about what you've watched. "Every time you use the platform, the algorithm is updated with new data so it can understand you more precisely," says Jake Moore, cybersecurity specialist at security company ESET.

#### TIKTOK

**How it works:** TikTok's algorithm uses machine learning to determine what content a user is most likely to engage with and serve them more of it, by finding videos that are similar or that are liked by people with similar user preferences.

- When users open TikTok for the first time, they are shown 8 popular videos featuring different trends, music, and topics. After that, the algorithm will continue to serve the user new iterations of 8 videos based on which videos the user engages with and what the user does.
- The algorithm identifies similar videos to those that have engaged a user based on video information, which could include details like captions, hashtags or sounds.
   Recommendations also take into account user device and account settings, which include data like language preference, country setting, and device type.
- Once TikTok collects enough data about the user, the app is able to map a user's preferences in relation to similar users and group them into "clusters." Simultaneously, it also groups videos into "clusters" based on similar themes, like "basketball" or "bunnies."
- **Using machine learning,** the algorithm serves videos to users based on their proximity to other clusters of users and content that they like.
- **TikTok's logic aims** to avoid redundancies that could bore the user, like seeing multiple videos with the same music or from the same creator.



#### TIKTOK

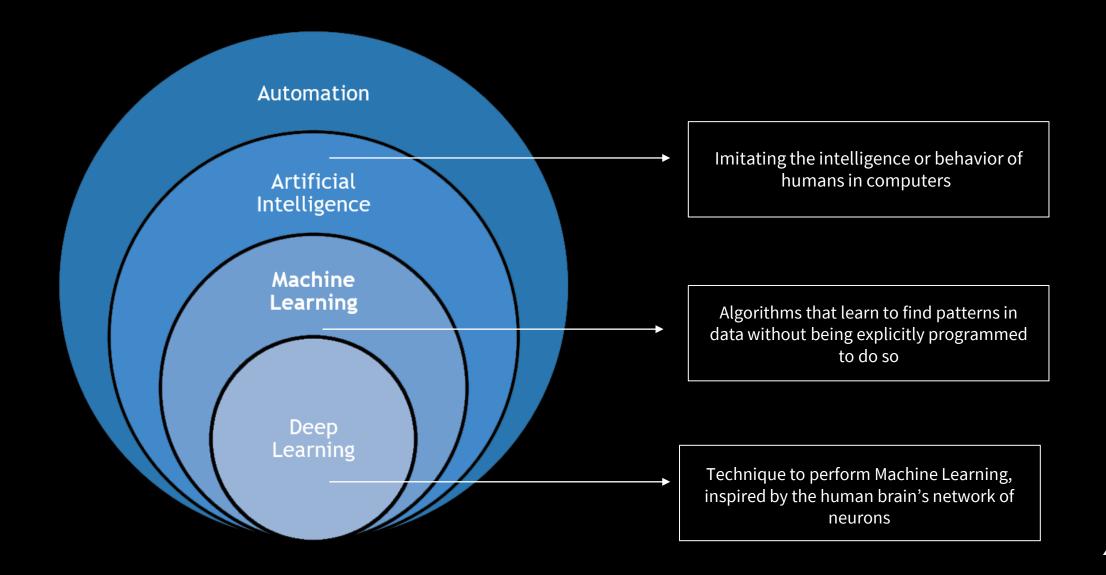
### What TikTok knows about you

TikTok can gather information when you arrive on the site even if you aren't signed up, via cookies and other trackers. Once you've created an account, the social network collects data about your activities and preferences based on the videos you watch.

TikTok knows the device you are using, your location, IP address, search history, the content of your messages, what you're viewing and for how long. It also collects device identifiers to track your interactions with advertisers. TikTok "infers" factors such as your age range, gender and interests based on the information it has about you. In the US, TikTok can <u>collect</u> biometric information including face and voiceprints.



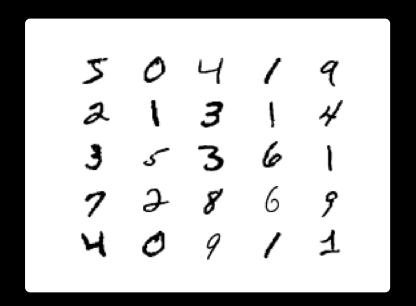
#### **OVERVIEW OF AI & ML**





#### HOW DO MACHINES LEARN?

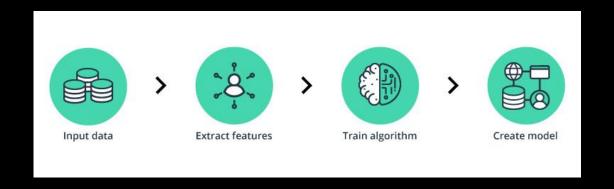
- Humans and computers learn in a similar fashion through experience
- Computers require vast
   amounts of data to be able to
   find trends and patterns





#### HOW DO MACHINES LEARN?

- Computers learn by being trained, by being fed data and being told what it is
- Computers utilize statistical techniques to find patterns in data that would be impossible to find by hand





- Suppose you had to write a computer program where given an image, you have to determine whether it is of a cat or a dog. What would you do?
- You would have to start by defining a set of rules that specify which attributes are unique to cats and dogs





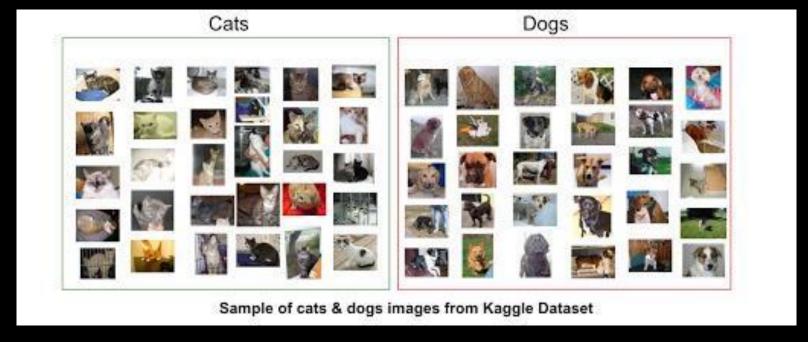


- How would your rules work for images where the full body of the animal is not shown? Or if there is another object in the image?
- For every rule you could make, there will always be images to bypass those rules because you did not account for something



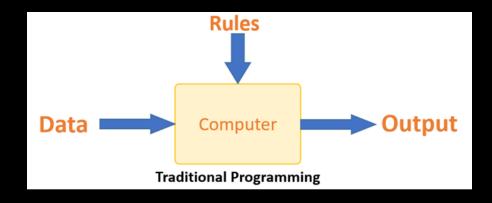


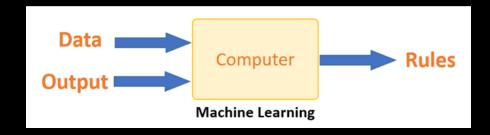
By feeding a machine learning algorithm thousands of images of cats and dogs, it will learn the rules itself and be able to classify new images with very high accuracy.



This is a common project in Computer Vision, we held a meeting on this topic last year, check out the recording <a href="here">here</a>

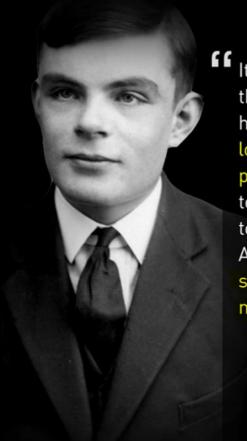
- In traditional programming, a developer manually writes all the rules for the data.
- In machine learning, the data and output are fed to the computer so it can be trained
- The computer learns how the data and output are correlated and generates rules based on those correlations







#### WILL MACHINES TAKE OVER?



It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers... They would be able to converse with each other to sharpen their wits.

At some stage therefore, we should have to expect the machines to take control.

- Alan Turing in 1951 For context, Turing conceived the idea of the modern computer in 1936 - known as the Turing Machine

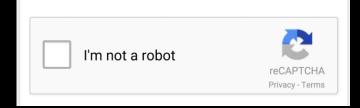
Turing Machines laid the groundwork for the creation of Computer Science and lead to the creation of all modern computational devices!

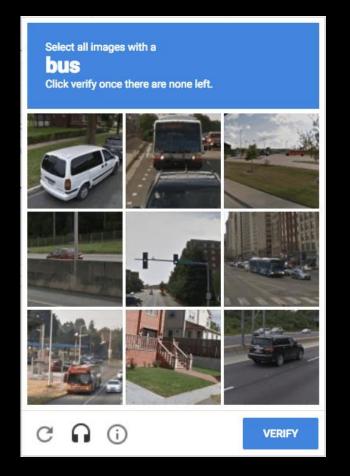


### **CAPTCHA**

Completely
Automated
Public
Turing Test to Tell
Computers and
Humans
Apart

Google uses CAPTCHA to train ML models for their self driving cars! Read more <u>here</u>









# How is breast cancer traditionally diagnosed?

- Breast ultrasound
- Diagnostic mammogram

- Breast magnetic resonance imaging (MRI)
- Biopsy

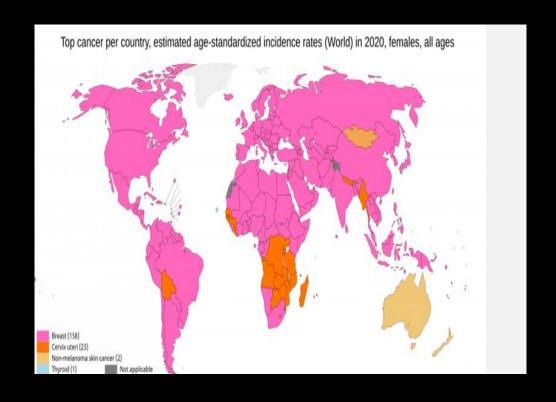
# What tests are traditionally used to assign stages to breast cancer?

- Blood tests (complete blood count)
- Mammograms of the other breast to look for indications of cancer
- Breast MRI
- CT scans
- PET scans



# **Breast Cancer Statistics Worldwide**

- About 1 in 18 (13% of the population) women will develop breast cancer.
- Recent 2020 studies have shown that more than 2.3 million women worldwide were diagnosed with a fatality rate of 685,000 (29.7%)
- As of the end of 2020 7.8 million women alive were diagnosed with breast cancer in the past 5 years, proving it to be the world's most prevalent cancer.
- Every 14 seconds a woman is diagnosed with breast cancer
- 5 year survival rate: 90%, 10 year survival rate: 84%,
   5 year survival rate(invasive found only in breast): 99%





#### This is a real breast cancer diagnostic data set

842302 M 842517 M 84300903 M	- CONTROL OF	us_me: texture 17.99	_mean perimete	or moanare						Constitution of the Consti		M	N	0	P Q	R	2	- 1	UV	W		Υ	2	AA	AB	AC	AD	AE	AF AG
842517 M 84300903 M	_	17.99														The state of the s			symmetry_fractal_dimra		And the second	the party of the second of the party of	and the second second						
84300903 M			10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471		0.07871	1.095	0.9053	8.589	153.4 0.006399					25.38	17.33	184.6	2019	0.1622	0.6656	0.7119	0.2654	0.4601	0.1189
THE RESERVE OF THE PARTY OF THE		20.57	17.77	132.9			0.07864	0.0869	0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08 0.005225		0.0186	0.0134		24.99	23.41	158.8	1956	0.1238	0.1866	0.2416	0.186	0.275	0.08902
		19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999	0.7456	0.7869	4.585	94.03 0.00615	-	0.03832	and promption of the control of	0.0225 0.004571	23.57	25.53	152.5	1709	0.1444	0.4245	0.4504	0.243	0.3613	0.08758
84348301 M		11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	0.4956	1.156	3.445	27.23 0.00911			0.01867	0.05963 0.009208	14.91	26.5	98.87	567.7	0.2098	0.8663	0.6869	0.2575	0.6638	0.173
84358402 M		20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883	0.7572	0.7813	5.438	94.44 0.01149	THE RESERVE OF THE PERSON NAMED IN	0.05688	0.01885	0.01756 0.005115	22.54	16.67	152.2	1575	0.1374	0.205	0.4	0.1625	0.2364	0.07678
843786 M		12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	0.3345	0.8902	2.217	27.19 0.00751	THE RESERVE AND ADDRESS.	-10000000000000000000000000000000000000	print believe with the land	0.02165 0.005082	15.47	23.75	103.4	741.6	0.1791	0.5249	0.5355	0.1741	0.3985	0.1244
844359 M		18.25	19.98	119.6		0.09463	0.109	0.1127	0.074	0.1794	0.05742	0.4467	0.7732	3.18	53.91 0.004314					22.88	27.66	153.2	1606	0.1442	0.2576	0.3784	0.1932	0.3063	0.08368
84458202 M		13.71	20.83	90.2	577.9	0.1189		0.09366	0.05985	0.2196	0.07451	0.5835	1.377	3.856		0.03029			- VINO 504111 NO 1000 NO 50 NO	17.06	28.14	110.6	897	0.1654	0.3682	0.2678	0.1556	0.3196	0.1151
0 844981 M		13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389	0.3063	1.002	2.406	24.32 0.005731				0.02143 0.003749	15.49	30.73	106.2	739.3	0.1703	0.5401	0.539	0.206	0.4378	0.1072
1 84501001 M		12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243	0.2976	1.599	2.039	23.94 0.007149			A STATE OF THE PARTY OF THE PAR	0.01789 0.01008	15.09	40.68	97.65	711.4	0.1853	1.058	1.105	0.221	0.4366	0.2075
2 845636 M		16.02	23.24	102.7		450000000000000000000000000000000000000		0.03299	0.03323	0.1528	0.05697	0.3795	1.187	2.466	40.51 0.004029				0.0146 0.003042	19.19	33.88	123.8	1150	0.1181	0.1551	0.1459	0.09975	0.2948	0.08452
3 84610002 M		15.78	17.89	103.6	781	0.0971	THE RESERVE TO STREET	0.09954	0.06606	0.1842	0.06082	0.5058	0.9849	3.564	54.16 0.005771		0.02791	0.01282	0.02008 0.004144	20.42	27.28	136.5	1299	0.1396	0.5609	0.3965	0.181	0.3792	0.1048
4 846226 M		19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078	0.9555	3.568	11.07	116.2 0.003139	Williams Street Co. St. St.	0.0889	0.0409	0.04484 0.01284	20.96	29.94	151.7	1332	0.1037	0.3903	0.3639	0.1767	0.3176	0.1023
5 846381 M		15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338	0.4033	1.078	2.903	36.58 0.009769			0.01992	0.02981 0.003002	16.84	27.66	112	876.5	0.1131	0.1924	0.2322	0.1119	0.2809	0.06287
6 84667401 M		13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069	0.07682	0.2121	1.169	2.061	19.21 0.006429			0.01628		15.03	32.01	108.8	697.7	0.1651	0.7725	0.6943	0.2208	0.3596	0.1431
7 84799002 M		14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077	0.37	1.033	2.879	32.55 0.005607	0.0424		0.0109	0.01857 0.005466	17.46	37.13	124.1	943.2	0.1678	0.6577	0.7026	0.1712	0.4218	0.1341
8 848406 M		14.68	20.13	94.74	684.5	0.09867		0.07395	0.05259	0.1586	0.05922	0.4727	1.24	3.195	45.4 0.005718				0.0141 0.002085	19.07	30.88	123.4	1138	0.1464	0.1871	0.2914	0.1609	0.3029	0.08216
9 84862001 M 0 849014 M	_	16.13	20.68	108.1	798.8 1260	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	0.5692	1.073	3.854 5.865	54.18 0.007026		0.03188		0.01689 0.004142	20.96	31.48	136.8 186.8	1315 2398	0.1789	0.4233	0.4784	0.2073	0.3706	0.1142
1 8510426 B		19.81	14.36	130 87.46				0.1479	0.09498	0.1582	0.05766	0.7582	0.7886	2.058	112.4 0.006494				0.01356 0.001997	15.11		99.7		0.1512	0.1773		0.2388	0.2768	0.07259
8510426 B 2 8510653 B		13.08	15.71	85.63	520	0.1075	and has properly and the	0.06664	0.04781	0.1883	0.05766	0.2699	0.7477	1.383	23.56 0.008462 14.67 0.004097	0.0146		0.01315	0.0198 0.0023 0.01678 0.002425	14.5	19.26 20.49	96.09	711.2 630.5	0.144	0.1773	0.239	0.1288	0.2977	0.07259
8510834 B		9.504	12.44	60.34	273.9			0.04368	0.0311	0.1967	0.06905	0.1832	0.9768	1.909	15.7 0.009606				0.02027 0.002968	10.23	15.66	65.13	314.9	0.1312	0.1148	0.08867	0.07283	0.245	0.08183
4 8511133 M		15.34	14.26	102.5	704.4	0.1024	0.2135	0.2077	0.02076	0.2521	0.07032	0.4388	0.7096	3.384	44.91 0.006789					18.07	19.08	125.1	980.9	0.1324	0.5954	0.6305	0.2393	0.4667	0.09946
5 851509 M		21.16	23.04	137.2	1404	0.09428	0.1022	0.1097	0.08632	0.1769	0.05278	0.6917	1.127	4.303	93.99 0.004728		0.00446		0.01083 0.001987	29.17	35.59	188	2615	0.1401	0.3934	0.3155	0.2009	0.2822	0.07526
6 852552 M		16.65	21.38	110	904.6	0.1121	0.1022	0.1525	0.0917	0.1995	0.0633	0.8068	0.9017	5.455	102.6 0.006048		0.02741	0.0113		26.46	31.56	177	2215	0.1805	0.3578	0.4695	0.2003	0.3613	0.09564
7 852631 M		17.14	16.4	116	912.7	0.1186	0.2276	0.2229	0.1401	0.304	0.07413	1.046	0.976	7.276		0.03799	0.03732		0.02308 0.007444	22.25	21.4	152.4	1461	0.1545	0.3949	0.3853	0.255	0.4066	0.1059
8 852763 M	-	14.58	21.53	97.41	644.8	0.1054	0.1868	0.1425	0.08783	0.2252	0.06924	0.2545	0.9832	2.11	21.05 0.004452		0.02681	0.01352	0.01454 0.003711	17.62	33.21	122.4	896.9	0.1525	0.6643	0.5539	0.2701	0.4264	0.1275
9 852781 M	-	18.61	20.25	122.1	1094	0.0944	0.1066	0.149	0.07731	0.1697	0.05699	0.8529	1.849	5.632	93.54 0.01075	0.02722	0.05081	0.01911	0.02293 0.004217	21.31	27.26	139.9	1403	0.1338	0.2117	0.3446	0.149	0.2341	0.07421
0 852973 M		15.3	25.27	102.4	732.4	0.1082	0.1697	0.1683	0.08751	0.1926	0.0654	0.439	1.012	3.498	43.5 0.005233				0.01768 0.002967	20.27	36.71	149.3	1269	0.1641	0.611	0.6335	0.2024	0.4027	0.09876
1 853201 M		17.57	15.05	115	955.1	0.09847	0.1157	0.09875	0.07953	0.1739	0.06149	0.6003	0.8225	4.655	61.1 0.005627	0.03033	0.03407	0.01354	0.01925 0.003742	20.01	19.52	134.9	1227	0.1255	0.2812	0.2489	0.1456	0.2756	0.07919
2 853401 M		18.63	25.11	124.8	1088	0.1064	0.1887	0.2319	0.1244	0.2183	0.06197	0.8307	1.466	5.574				0.01158		23.15	34.01	160.5	1670	0.1491	0.4257	0.6133	0.1848	0.3444	0.09782
853612 M	_	11.84	18.7	77.93	440.6	0.1109	0.1516	0.1218	0.05182	0.2301	0.07799	0.4825	1.03	3.475	41 0.005551	0.03414				16.82	28.12	119.4	888.7	0.1637	0.5775	0.6956	0.1546	0.4761	0.1402
4 85382601 M	_	17.02	23.98	112.8	899.3	0.1197	0.1496	0.2417	0.1203	0.2248	0.06382	0.6009	1.398	3.999	67,78 0.008268	0.03082				20.88	32.09	136.1	1344	0.1634	0.3559	0.5588	0.1847	0.353	0.08482
5 854002 M		19.27	26.47	127.9	1162	0.09401	0.1719	0.1657	0.07593	0.1853	0.06261	0.5558	0.6062	3.528	68.17 0.005015				0.01543 0.003896	24.15	30.9	161.4	1813	0.1509	0.659	0.6091	0.1785	0.3672	0.1123
6 854039 M	1	16.13	17.88	107	807.2	0.104	0.1559	0.1354	0.07752	0.1998	0.06515	0.334	0.6857	2.183	35.03 0.004185	A CONTRACTOR OF THE	100000000000000000000000000000000000000	A PROPERTY OF THE PARTY OF	0.01703 0.003817	20.21	27.26	132.7	1261	0.1446	0.5804	0.5274	0.1864	0.427	0.1233
7 854253 M		16.74	21.59	110.1	869.5	0.0961	0.1336	0.1348	0.06018	0.1896	0.05656	0.4615	0.9197	3.008	45.19 0.005776					20.01	29.02	133.5	1229	0.1563	0.3835	0.5409	0.1813	0.4863	0.08633
8 854268 M	_	14.25	21.72	93.63	and the same of the same of	0.09823	0.1098	0.1319	0.05598	0.1885	0.06125	0.286	1.019	2.657	24.91 0.005878	AND RESIDENCE OF THE PARTY OF	0.04815		0.02028 0.004022	15.89	30.36	116.2	799.6	0.1446	0.4238	0.5186	0.1447	0.3591	0.1014
9 854941 B	_	13.03	18.42	82.61				0.02562	0.02923	0.1467	0.05863	0.1839	2.342	1.17	14.16 0.004352			0.01164		13.3	22.81	84.46		0.09701	0.04619	0.04833	0.05013	0.1987	0.06169
0 855133 M		14.99	25.2	95.54		Professional Profe	0.000	0.02398	0.02899	0.1565	0.05504	1.214	2.188	8.077	106 0.006883				0.007882 0.001754	14.99	25.2	95.54	-37	0.09387		0.02398	0.02899	0.1565	0.05504
1 855138 M		13.48	20.82	88.4	559.2	0.1016	0.1255	0.1063	0.05439	0.172	0.06419	0.213	0.5914	1.545	18.52 0.005367	0.02239		0.01262		15.53	26.02	107.3	740.4	0.161	0.4225	0.503	0.2258	0.2807	0.1071
2 855167 M		13.44	21.58	86.18			0.06031	0.0311	0.02031	0.1784	0.05587	0.2385	0.8265	1.572	20.53 0.00328			0.006881	0.0138 0.001286	15.93	30.25	102.5	787.9	0.1094	0.2043	0.2085	0.1112	0.2994	0.07146
855563 M		10.95	21.35	71.9	371.1	0.1227	0.1218	0.1044	0.05669	0.1895	0.0687	0.2366	1.428	1.822		0.01764			0.01357 0.00304	12.84	35.34	87.22	514	0.1909	0.2698	0.4023	0.1424	0.2964	0.09606
4 855625 M		19.07	24.81	128.3		0.09081	0.219	0.2107	0.09961	0.231	0.06343	0.9811	1.666	8.83	104.9 0.006548			0.02638		24.09	33.17	177.4	1651	0.1247	0.7444	0.7242	0.2493	0.467	0.1038
5 856106 M	_	13.28	20.28	87.32	545.2	0.1041		0.09847	0.06158	0.1974	0.06782	0.3704	0.8249	2.427	31.33 0.005072					17.38	28	113.1	907.2	0.153	0.3724	0.3664	0.1492	0.3739	0.1027
6 85638502 M		13.17	21.81	85.42		0.09714		0.08259			0.06177	0.1938	0.6123	1.334					0.01113 0.00172	16.23	29.89	105.5	740.7	0.1503	0.3904	0.3728	0.1607	0.3693	0.09618

How would you determine the factors that lead to the diagnosis being malignant or benign?

- A multitude of if statements?



#### Features

(Radius, Texture, Concavity, Area, Perimeter, Smoothness, Compactness, etc.)

Diagnosis:

Malignant or Benign

(Labels)



#### LAB INSTRUCTIONS

Go to <a href="https://t.ly/PLwd">https://t.ly/PLwd</a>

Check out the completed code <u>here</u>

Follow along on Google Colab

- Google Colab is a cloud platform where we can run Python code
  - O Python is the most popular programming language for ML

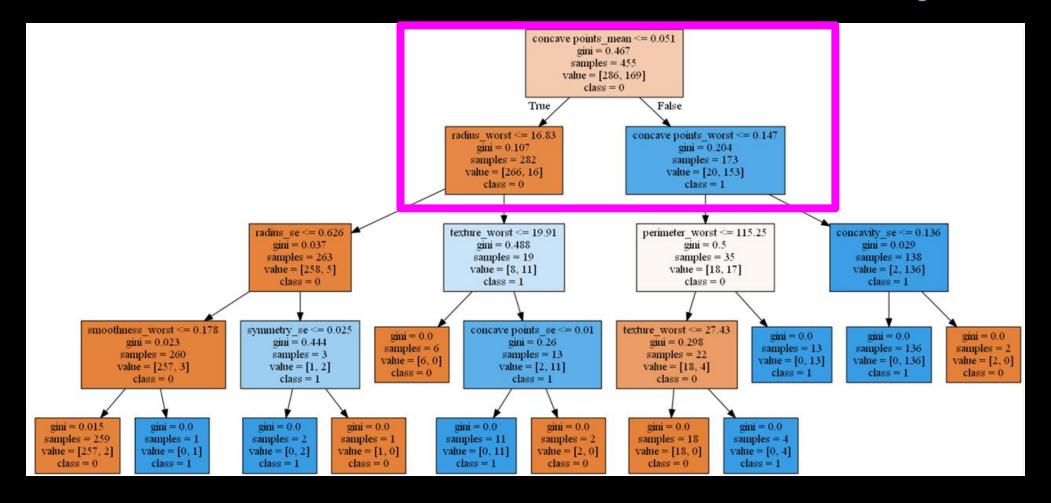


#### DECISION TREE

Rules the Computer Learned from the Data

Orange: Benign

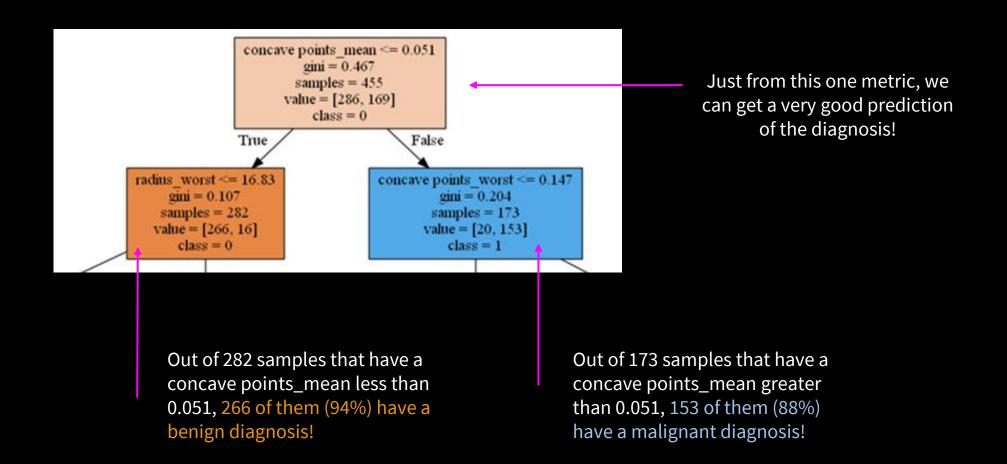
Blue: Malignant





#### **DECISION TREE**

This would've been unfeasible to find manually without the use of ML!



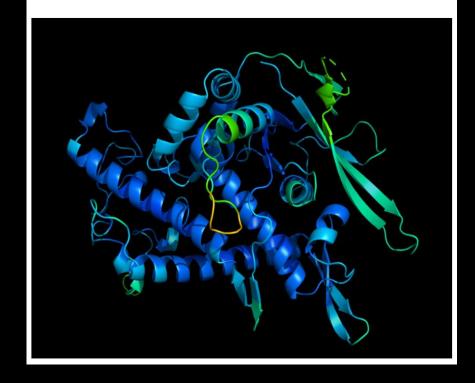


#### AI IS REVOLUTIONIZING THE MEDICAL INDUSTRY

NEWS · 30 NOVEMBER 2020

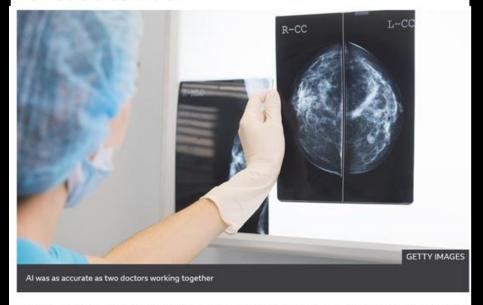
## 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.





## AI 'outperforms' doctors diagnosing breast cancer



An international team, including researchers from **Google Health** and **Imperial College London**, designed and trained a computer model on X-ray images from nearly 29,000 women.

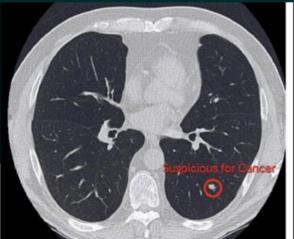
The algorithm outperformed six radiologists in reading mammograms.

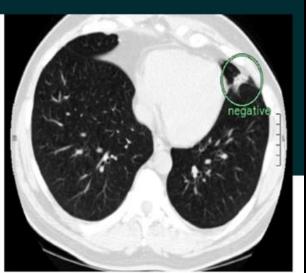
AI was still as good as two doctors working together.

Unlike humans, AI is tireless. Experts say it could improve detection.

## Google's lung cancer detection AI outperforms 6 human radiologists

Windshipolinson May 20, 2019 0.00 A





Google AI researchers working with Northwestern Medicine created an AI model capable of detecting lung cancer from screening tests better than human radiologists with an average of eight years experience.

#### AUTOMATED COVID-19 DETECTION WITH ML

Check out the source code <u>here</u>





We taught computers how to learn,

Now computers are helping us learn.

#### Additional Resources

- Stanford HAI: The Future of Artificial
   Intelligence in Medicine and Imaging
- How AI & Neuroscience Drive Each other
   Forwards
- How Machine Learning is Transforming
   Healthcare
- Google AI: Advancing Healthcare Research & AI in Medicine

