

# Métodos Numéricos

Departamento de Computación  
Facultad de Ciencias Exactas y Naturales  
Universidad de Buenos Aires

28 de Octubre de 2016



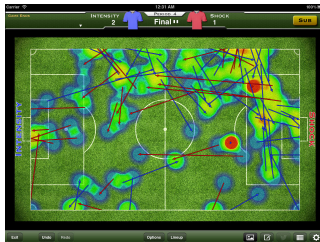
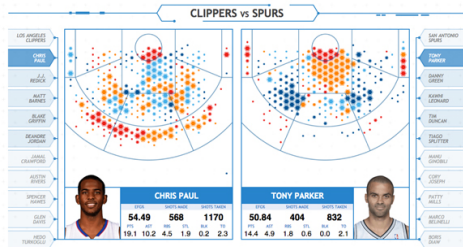
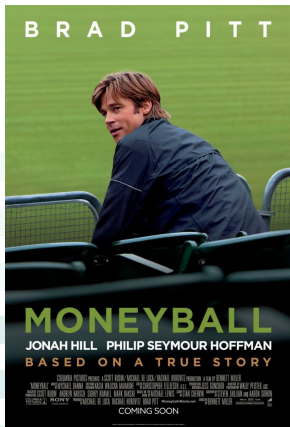
# Hasta ahora...

- ▶ Aproximación de temperaturas en un alto horno
- ▶ Reconstrucción de imágenes eliminando ruido
- ▶ Reconocimiento facial
- ▶ Clasificación de noticias
- ▶ PageRank
- ▶ Generación de imágenes de tomografía computada
- ▶ Hoy: predicción de desempeño de equipos en NBA

# Motivación

## Sport Analytics

Utilización de técnicas estadísticas, inteligencia artificial y optimización en el deporte.



# Orígenes de sport analytics

- ▶ En MLB surge la idea de usar programas estadísticos en los 70/80 y en 1980 surge el término *sabermetrics*.
- ▶ En NFL las estadísticas comienzan a usarse activamente en los 90/2000.
- ▶ ¿Qué pasa en el fútbol? Sólo algunos clubes de las principales ligas declararon haber comenzado a hacer uso de modelos matemáticos, principalmente en Inglaterra e Italia en los últimos años.
- ▶ En la NBA, datos de cuya liga analizaremos en este trabajo, históricamente se relevaban estadísticas pero sólo en los 90/2000 se comenzaron a diseñar indicadores específicos que contemplaran la dinámica del juego.

# Orígenes de sport analytics

- ▶ En MLB surge la idea de usar programas estadísticos en los 70/80 y en 1980 surge el término *sabermetrics*.
- ▶ En NFL las estadísticas comienzan a usarse activamente en los 90/2000.
- ▶ ¿Qué pasa en el fútbol? Sólo algunos clubes de las principales ligas declararon haber comenzado a hacer uso de modelos matemáticos, principalmente en Inglaterra e Italia en los últimos años.
- ▶ En la NBA, datos de cuya liga analizaremos en este trabajo, históricamente se relevaban estadísticas pero sólo en los 90/2000 se comenzaron a diseñar indicadores específicos que contemplaran la dinámica del juego.

# Orígenes de sport analytics

- ▶ En MLB surge la idea de usar programas estadísticos en los 70/80 y en 1980 surge el término *sabermetrics*.
- ▶ En NFL las estadísticas comienzan a usarse activamente en los 90/2000.
- ▶ ¿Qué pasa en el fútbol? Sólo algunos clubes de las principales ligas declararon haber comenzado a hacer uso de modelos matemáticos, principalmente en Inglaterra e Italia en los últimos años.
- ▶ En la NBA, datos de cuya liga analizaremos en este trabajo, históricamente se relevaban estadísticas pero sólo en los 90/2000 se comenzaron a diseñar indicadores específicos que contemplaran la dinámica del juego.

# Orígenes de sport analytics

- ▶ En MLB surge la idea de usar programas estadísticos en los 70/80 y en 1980 surge el término *sabermetrics*.
- ▶ En NFL las estadísticas comienzan a usarse activamente en los 90/2000.
- ▶ ¿Qué pasa en el fútbol? Sólo algunos clubes de las principales ligas declararon haber comenzado a hacer uso de modelos matemáticos, principalmente en Inglaterra e Italia en los últimos años.
- ▶ En la NBA, datos de cuya liga analizaremos en este trabajo, históricamente se relevaban estadísticas pero sólo en los 90/2000 se comenzaron a diseñar indicadores específicos que contemplaran la dinámica del juego.

# Orígenes de sport analytics

- ▶ En MLB surge la idea de usar programas estadísticos en los 70/80 y en 1980 surge el término *sabermetrics*.
- ▶ En NFL las estadísticas comienzan a usarse activamente en los 90/2000.
- ▶ ¿Qué pasa en el fútbol? Sólo algunos clubes de las principales ligas declararon haber comenzado a hacer uso de modelos matemáticos, principalmente en Inglaterra e Italia en los últimos años.
- ▶ En la NBA, datos de cuya liga analizaremos en este trabajo, históricamente se relevaban estadísticas pero sólo en los 90/2000 se comenzaron a diseñar indicadores específicos que contemplaran la dinámica del juego.



# Orígenes de sport analytics

- ▶ En MLB surge la idea de usar programas estadísticos en los 70/80 y en 1980 surge el término *sabermetrics*.
- ▶ En NFL las estadísticas comienzan a usarse activamente en los 90/2000.
- ▶ ¿Qué pasa en el fútbol? Sólo algunos clubes de las principales ligas declararon haber comenzado a hacer uso de modelos matemáticos, principalmente en Inglaterra e Italia en los últimos años.
- ▶ En la NBA, datos de cuya liga analizaremos en este trabajo, históricamente se relevaban estadísticas pero sólo en los 90/2000 se comenzaron a diseñar indicadores específicos que contemplaran la dinámica del juego.

# ¿Analytics?, ¿para qué?

También Big Data o Data Science

1. Observar los datos
2. Analizarlos y entenderlos
3. Tomar decisiones

En este TP: plantear métricas sobre el rendimiento de los equipos y jugadores para predecir el desempeño del equipo al finalizar la temporada regular.

# ¿Analytics?, ¿para qué?

También Big Data o Data Science

1. Observar los datos
2. Analizarlos y entenderlos
3. Tomar decisiones

En este TP: plantear métricas sobre el rendimiento de los equipos y jugadores para predecir el desempeño del equipo al finalizar la temporada regular.

# Dominio del problema

## NBA y temporada regular

- ▶ 30 equipos, 15 en el Este y 15 en el Oeste
- ▶ Cada equipo juega 82 partidos
- ▶ No existe el empate

1

---

<sup>1</sup>Tomado de [http://www.basketball-reference.com/leagues/NBA\\_2016.html](http://www.basketball-reference.com/leagues/NBA_2016.html)

# Dominio del problema

## NBA y temporada regular

- ▶ 30 equipos, 15 en el Este y 15 en el Oeste
- ▶ Cada equipo juega 82 partidos
- ▶ No existe el empate

Eastern Conference		W	L	W/L%	GB	PS/G	PA/G	SRS
<a href="#">Cleveland Cavaliers</a> *	(1)	57	25	.695	—	104.3	98.3	5.45
<a href="#">Toronto Raptors</a> *	(2)	56	26	.683	1.0	102.7	98.2	4.08
<a href="#">Miami Heat</a> *	(3)	48	34	.585	9.0	100.0	98.4	1.50
<a href="#">Atlanta Hawks</a> *	(4)	48	34	.585	9.0	102.8	99.2	3.49
<a href="#">Boston Celtics</a> *	(5)	48	34	.585	9.0	105.7	102.5	2.84
<a href="#">Charlotte Hornets</a> *	(6)	48	34	.585	9.0	103.4	100.7	2.36
<a href="#">Indiana Pacers</a> *	(7)	45	37	.549	12.0	102.2	100.5	1.62
<a href="#">Detroit Pistons</a> *	(8)	44	38	.537	13.0	102.0	101.4	0.43
<a href="#">Chicago Bulls</a>	(9)	42	40	.512	15.0	101.6	103.1	-1.46
<a href="#">Washington Wizards</a>	(10)	41	41	.500	16.0	104.1	104.6	-0.50
<a href="#">Orlando Magic</a>	(11)	35	47	.427	22.0	102.1	103.7	-1.68
<a href="#">Milwaukee Bucks</a>	(12)	33	49	.402	24.0	99.0	103.2	-3.98
<a href="#">New York Knicks</a>	(13)	32	50	.390	25.0	98.4	101.1	-2.74
<a href="#">Brooklyn Nets</a>	(14)	21	61	.256	36.0	98.6	106.0	-7.12
<a href="#">Philadelphia 76ers</a>	(15)	10	72	.122	47.0	97.4	107.6	-9.92

Western Conference		W	L	W/L%	GB	PS/G	PA/G	SRS
<a href="#">Golden State Warriors</a> *	(1)	73	9	.890	—	114.9	104.1	10.38
<a href="#">San Antonio Spurs</a> *	(2)	67	15	.817	6.0	103.5	92.9	10.28
<a href="#">Oklahoma City Thunder</a> *	(3)	55	27	.671	18.0	110.2	102.9	7.09
<a href="#">Los Angeles Clippers</a> *	(4)	53	29	.646	20.0	104.5	100.2	4.13
<a href="#">Portland Trail Blazers</a> *	(5)	44	38	.537	29.0	105.1	104.3	0.98
<a href="#">Dallas Mavericks</a> *	(6)	42	40	.512	31.0	102.3	102.6	-0.02
<a href="#">Memphis Grizzlies</a> *	(7)	42	40	.512	31.0	99.1	101.3	-2.14
<a href="#">Houston Rockets</a> *	(8)	41	41	.500	32.0	106.5	106.4	0.34
<a href="#">Utah Jazz</a>	(9)	40	42	.488	33.0	97.7	95.9	1.84
<a href="#">Sacramento Kings</a>	(10)	33	49	.402	40.0	106.6	109.1	-2.32
<a href="#">Denver Nuggets</a>	(10)	33	49	.402	40.0	101.9	105.0	-2.81
<a href="#">New Orleans Pelicans</a>	(12)	30	52	.366	43.0	102.7	106.5	-3.56
<a href="#">Minnesota Timberwolves</a>	(13)	29	53	.354	44.0	102.4	106.0	-3.38
<a href="#">Phoenix Suns</a>	(14)	23	59	.280	50.0	100.9	107.5	-6.32
<a href="#">Los Angeles Lakers</a>	(15)	17	65	.207	56.0	97.3	106.9	-8.92

<sup>1</sup>Tomado de [http://www.basketball-reference.com/leagues/NBA\\_2016.html](http://www.basketball-reference.com/leagues/NBA_2016.html)

# Datos

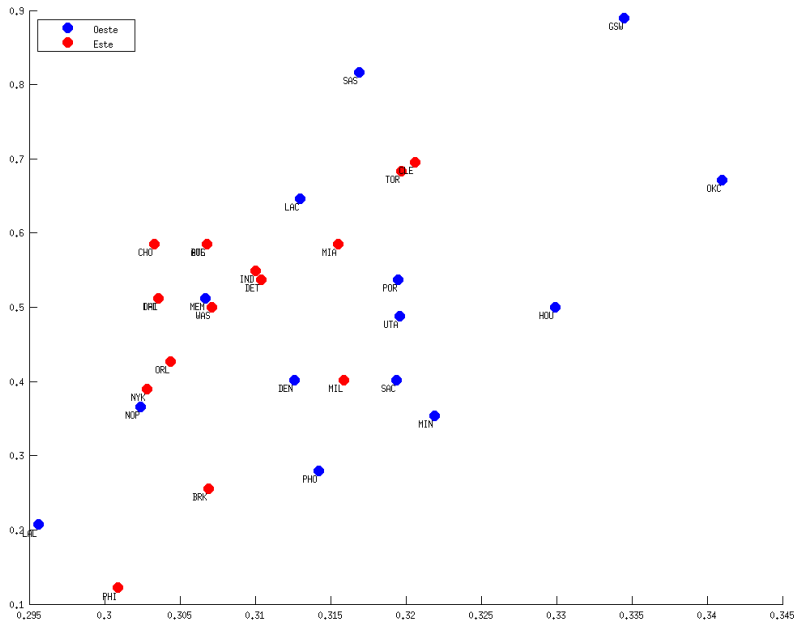
Rk	Team	G	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	PTS/G
1	<a href="#">Golden State Warriors*</a>	82	19880	3489	7159	.487	1077	2592	.416	2412	4567	.528	1366	1790	.763	816	2972	3788	2373	689	498	1245	1701	9421	114.9
2	<a href="#">Oklahoma City Thunder*</a>	82	19830	3372	7082	.476	678	1945	.349	2694	5137	.524	1616	2067	.782	1071	2916	3987	1883	603	487	1305	1691	9038	110.2
3	<a href="#">Sacramento Kings</a>	82	19805	3283	7083	.464	660	1839	.359	2623	5244	.500	1514	2089	.725	868	2760	3628	2009	733	368	1326	1676	8740	106.6
4	<a href="#">Houston Rockets*</a>	82	19830	3094	6847	.452	878	2533	.347	2216	4314	.514	1671	2407	.694	930	2601	3531	1821	821	430	1307	1790	8737	106.5
5	<a href="#">Boston Celtics*</a>	82	19780	3216	7318	.439	717	2142	.335	2499	5176	.483	1520	1929	.788	950	2733	3683	1981	752	348	1127	1796	8669	105.7
6	<a href="#">Portland Trail Blazers*</a>	82	19805	3167	7040	.450	864	2336	.370	2303	4704	.490	1424	1889	.754	948	2782	3730	1748	562	380	1200	1782	8622	105.1
7	<a href="#">Los Angeles Clippers*</a>	82	19830	3141	6759	.465	797	2190	.364	2344	4569	.513	1490	2152	.692	721	2727	3448	1873	709	460	1063	1746	8569	104.5
8	<a href="#">Cleveland Cavaliers*</a>	82	19855	3171	6888	.460	880	2428	.362	2291	4460	.514	1333	1783	.748	873	2777	3650	1861	551	317	1114	1666	8555	104.3
9	<a href="#">Washington Wizards</a>	82	19755	3238	7033	.460	709	1983	.358	2529	5050	.501	1349	1849	.730	743	2688	3431	2005	708	323	1186	1708	8534	104.1
10	<a href="#">San Antonio Spurs*</a>	82	19705	3289	6797	.484	570	1518	.375	2719	5279	.515	1342	1672	.803	770	2831	3601	2010	677	485	1071	1433	8490	103.5
11	<a href="#">Charlotte Hornets*</a>	82	19855	3036	6922	.439	873	2410	.362	2163	4512	.479	1534	1941	.790	734	2869	3603	1778	595	438	1030	1487	8479	103.4
12	<a href="#">Atlanta Hawks*</a>	82	19830	3168	6923	.458	815	2326	.350	2353	4597	.512	1282	1638	.783	679	2772	3451	2100	747	486	1226	1570	8433	102.8

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	<a href="#">Quincy Acy</a>	PF	25	<a href="#">SAC</a>	59	29	876	119	214	.556	19	49	.388	100	165	.606	.600	50	68	.735	65	123	188	27	29	24	27	103	307
2	<a href="#">Jordan Adams</a>	SG	21	<a href="#">MEM</a>	2	0	15	2	6	.333	0	1	.000	2	5	.400	.333	3	5	.600	0	2	2	3	3	0	2	2	7
3	<a href="#">Steven Adams</a>	C	22	<a href="#">OKC</a>	80	80	2014	261	426	.613	0	0		261	426	.613	.613	114	196	.582	219	314	533	62	42	89	84	223	636
4	<a href="#">Arron Affalo</a>	SG	30	<a href="#">NYK</a>	71	57	2371	354	799	.443	91	238	.382	263	561	.469	.500	110	131	.840	23	243	266	144	25	10	82	142	909
5	<a href="#">Alexis Ajinca</a>	C	27	<a href="#">NOP</a>	59	17	861	150	315	.476	0	1	.000	150	314	.478	.476	52	62	.839	75	194	269	31	19	36	54	134	352
6	<a href="#">Cole Aldrich</a>	C	27	<a href="#">LAC</a>	60	5	800	134	225	.596	0	0		134	225	.596	.596	60	84	.714	86	202	288	50	47	68	64	139	328
7	<a href="#">LaMarcus Aldridge</a>	PF	30	<a href="#">SAS</a>	74	74	2261	536	1045	.513	0	16	.000	536	1029	.521	.513	259	302	.858	176	456	632	110	38	81	99	151	1331
8	<a href="#">Cliff Alexander</a>	PF	20	<a href="#">POR</a>	8	0	36	5	10	.500	0	0		5	10	.500	.500	0	0		2	4	6	0	1	2	1	1	10
9	<a href="#">Lavoy Allen</a>	PF	26	<a href="#">IND</a>	79	28	1599	191	370	.516	0	0		191	370	.516	.516	46	73	.630	162	262	424	76	26	42	69	147	428
10	<a href="#">Tony Allen</a>	SG	34	<a href="#">MEM</a>	64	57	1620	215	469	.458	15	42	.357	200	427	.468	.474	90	138	.652	104	192	296	70	110	18	78	175	535
11	<a href="#">Al-Farouq Aminu</a>	SF	25	<a href="#">POR</a>	82	82	2341	299	719	.416	126	349	.361	173	370	.468	.503	115	156	.737	98	401	499	138	72	53	120	171	839
12	<a href="#">Lou Amundson</a>	PF	33	<a href="#">NYK</a>	29	0	204	19	53	.358	0	0		19	53	.358	.358	14	27	.519	23	26	49	12	6	7	15	44	52

## TP3: El problema

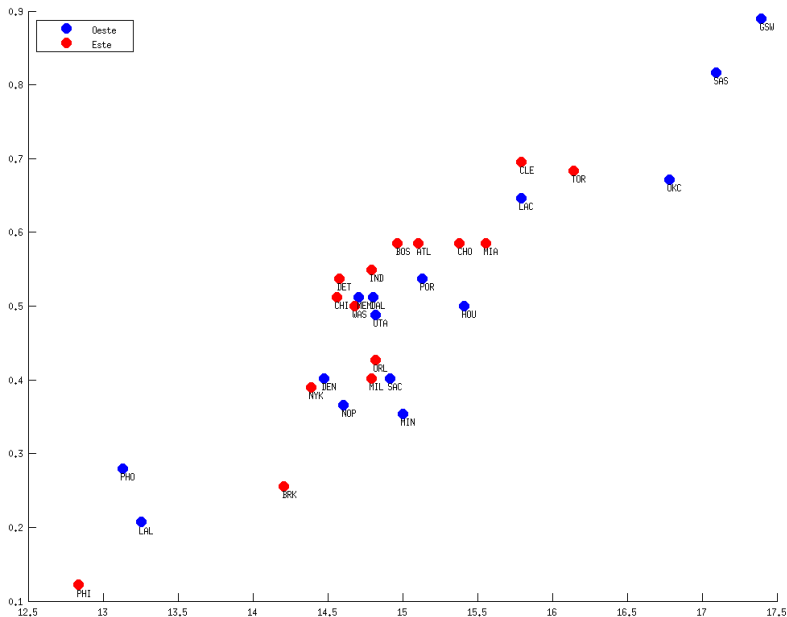
- ▶ Utilizar Cuadrados Mínimos Lineales (CML) como técnica e identificar modelos que describan el win rate (cantidad de partidos ganados sobre el total) a partir de los datos.
- ▶ Plantear dos métricas, una basada en estadísticas a nivel equipo y otra a nivel jugadores.
- ▶ Analizar los resultados y evaluarlos usando cross validation. Compararlos con Four Factors y Player Efficiency Rate.
- ▶ Utilizar los modelos para predecir el desempeño de los equipos.
- ▶ Aplicar las técnicas y metodologías aprendidas durante la materia.
- ▶ Libertad para plantear las métricas pero rigurosidad en su justificación y análisis.

# Four Factors vs Win Rate

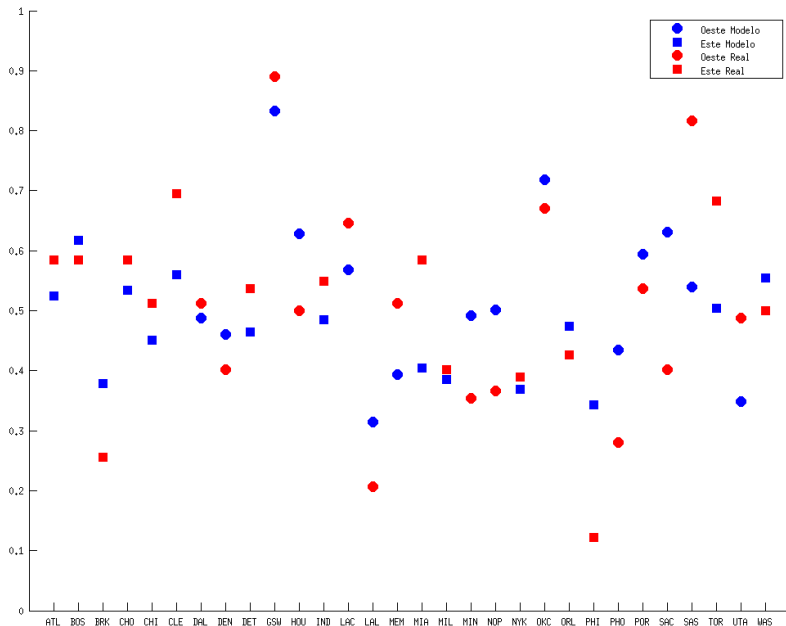




# Player Efficiency Rate vs Win Rate



# Modelo usando puntos a favor y en contra



# Usando CML

- ▶ Considerar todas las estadísticas disponibles y elegir las con criterio
- ▶ Considerar no sólo los datos sino eventualmente funciones aplicadas sobre los mismos (por ejemplo polinomios)
- ▶ Considerar datos de distintas temporadas y evaluar sobre temporadas futuras

# Métricas de evaluación (1/2)

- ▶  $N$  observaciones  $(x_{(i)}, y_{(i)})$ , con  $x_{(i)} \in \mathbb{R}^k$  el vector de *features* e  $y_{(i)} \in \mathbb{R}$  nuestra variable dependiente.
- ▶ Suponemos  $y_{(i)} = f(x_{(i)}) + \epsilon_i$ ,  $i = 1, \dots, N$ , donde  $\epsilon_i$  es el error de la medición  $i$ -ésima.
- ▶ Dado un modelo  $\hat{f}$  de  $f$  y  $(x_{(i)}, y_{(i)})$ , definimos  $\hat{y}_{(i)} = \hat{f}(x_{(i)})$  y  $e_{(i)} = y_{(i)} - \hat{y}_{(i)}$ . Con estas definiciones, podemos calcular el MSE del modelo  $\hat{f}$  como

$$MSE(\hat{f}) = \frac{1}{N} \sum_{i=1}^N e_{(i)}^2.$$

- ▶ Para evaluar cómo se comporta como modelo predictivo, podemos usar *Cross-Validation* combinado con MSE:

## Métricas de evaluación (2/2)

- ▶ Es posible que existan dependencias temporales en los datos.
- ▶ Para ello, consideramos que cada observación está asociada a un determinado período de tiempo  $t$ , con  $t = 1, \dots, T$ ,  $(x_{(i)}^t, y_{(i)}^t)$ , y asumimos que al menos  $K$  períodos de tiempo son necesarios para poder conformar el conjunto de *training*. Para evaluar los resultados de la predicción en el período  $\tau \in [K, T]$  se puede:
  1. Tomar los conjuntos de observaciones correspondientes a períodos  $1, \dots, \tau - 1$  como training.
  2. Calcular las métricas correspondientes tomando como test el período  $\tau$ .
  3. Al finalizar, reportar alguna medida sobre los resultados parciales obtenidos.

# Desarrollo

- ▶ El TP se puede hacer en MATLAB, Python y/o C++. Se pueden usar bibliotecas con los métodos implementados.
- ▶ Junto al enunciado se presentan scripts en bash que extraen features y la información necesaria para obtener PER y Four Factors. Estos scripts pueden resultar útiles como inspiración para plantear sus propias herramientas. Éstas pueden ser compartidas entre ustedes, siempre y cuando se mantenga reservada la información de los experimentos. Deben hacerlo a través de la lista de alumnos.

Machete: mostrar cómo funcionan

# Desarrollo

- ▶ El TP se puede hacer en MATLAB, Python y/o C++. Se pueden usar bibliotecas con los métodos implementados.
- ▶ Junto al enunciado se presentan scripts en bash que extraen features y la información necesaria para obtener PER y Four Factors. Estos scripts pueden resultar útiles como inspiración para plantear sus propias herramientas. Éstas pueden ser compartidas entre ustedes, siempre y cuando se mantenga reservada la información de los experimentos. Deben hacerlo a través de la lista de alumnos.
- ▶ Machete: mostrar cómo funcionan

# Presentación

- ▶ Además del informe usual (esta vez bajo otro formato), el trabajo será expuesto en una presentación oral frente a alumnos y docentes.
- ▶ Para ello, harán una presentación mostrando lo que hicieron. Vamos a dar soporte y ayuda para prepararla.
- ▶ La exposición contará con 30 minutos totales: 15 minutos para presentar, 15 minutos de preguntas y respuestas. La nota de aprobación es individual.
- ▶ Para poder presentar en primera fecha, se deberá tener la aprobación previa por parte de los docentes correctores.



# Trabajo Práctico

Fecha de entrega

- ▶ *Formato Electrónico*: Miércoles 16/11, hasta las 23:59 hs.
- ▶ *Confirmación presentación oral*: Viernes 18/11, por correo electrónico.
- ▶ *Presentación oral*: Lunes 21/11, en horario a determinar luego de la confirmación. Será en horario de clase de la materia.

## Importante

El horario es estricto. Los correos recibidos después de la hora indicada serán considerados re-entrega.