High Accuracy Forecasting of No Show Passengers

(Maximization of profit by correct overbooking)

Dr. Viterbo H. Berberena González

Director de Minería de Datos

Pearson S.A. de C.V.

Agenda

- El problema de la sobreventa de boletos en la aerolíneas.
- Los métodos clásicos de pronósticos y sus limitaciones.
- El enfoque de las técnicas y herramientas de minería de datos.
- El problema de la sobreventa de boletos en AeroMéxico.
- Los pasos fundamentales para la construcción del modelo predictivo.
- Los principales resultados.

The overbooking problem in airlines companies

- Voluntary denied boarding (volunteers to surrender their seats in exchange for advantages)
- Involuntary denied boarding (against a passenger's will)
  - Denied boarding compensation (European Union).
    - €250 for flights of less than 1500 km.
    - €400 for intra-Community flights of more than 1500 km and for other flights 1500 and 3500 km.
    - €600 for all other flights.
    - In addition to financial compensation, passengers denied boarding will continue to enjoy these rights: the choice between reimbursement of their ticket and an alternative flight, and meals, refreshments and hotel accommodation.
- Denied boarding causes passengers great inconvenience and loss of time.

Consumer Complaints Against Top U.S. Airlines by Category

Complaint category	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
TOTAL	**7,703**	**6,106**	**5,639**	**4,438**	**5,179**	**4,629**	**5,782**	**6,394**	**7,994**	**17,381**	**20,564**	**16,508**	**9,471**
Flight problems[1]	3,034	1,877	1,624	1,211	1,586	1,133	1,628	1,699	2,277	6,469	8,698	5,480	2,031
Customer service[2]	758	714	695	599	805	667	999	1,418	1,715	3,664	4,074	2,860	1,715
Baggage	1,329	883	752	627	761	628	882	826	1,108	2,353	2,753	2,490	1,421
Reservations/ticketing/ boarding[3]	624	659	680	577	598	666	857	904	1,137	1,328	1,405	1,611	1,159
Refunds	701	783	721	482	393	576	521	531	602	940	803	1,347	1,106
Oversales[4]	399	301	265	257	301	263	353	414	388	673	759	638	454
Fares[5]	312	388	573	398	267	185	180	195	277	584	708	666	523
Disability[3]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	526	612	508	477
Advertising	96	96	54	51	94	66	61	57	40	57	42	61	68
Tours	29	23	12	16	127	18	16	13	23	28	25	n.a.	n.a.
Smoking[6]	74	30	25	30	20	15	13	5	4	n.a.	n.a.	n.a.	n.a.
Credit[6]	5	10	10	4	2	4	3	1	1	n.a.	n.a.	n.a.	n.a.
Other[6]	342	342	228	186	225	408	269	331	422	759	675	650	322
Animals[7]	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0	1	6	0

Compensations as outlined by the US Department of Transportation

Domestic Travel

If the scheduled arrival time on the re-accommodated flight at the destination, compared to the scheduled arrival time on the original flight, is later by:

- One hour or less No Compensation is provided.
- Between one and two hours 100% of coupon value to first stopover point not to exceed $200 USD is provided.
- More than two hours 200% of coupon value to first stopover point not to exceed $400 USD.

International Travel

If the scheduled arrival on the re-accommodated flight at the destination, compared to the scheduled arrival time on the original flight is later by:

- Between one and four hours 100% of coupon value to first stopover point not to exceed $200 USD is provided.
- More than four hours 200% of coupon value to first stopover point not to exceed $400 USD is provided.

Passengers Denied Boarding by Top U.S. Airlines,[1] 2002

Rank Airline	Denied boardings (DBs)		Enplaned passengers	Involuntary DBs per 10,000 passengers
	Voluntary	Involuntary		
1.American Eagle	1,103	19	1,001,798	0.19
2.America West	52,593	385	19,711,035	0.20
3.American	135,989	2,650	86,792,674	0.31
4.U.S. Airways	101,084	1,526	43,978,481	0.35
5.Northwest	76,878	2,809	46,993,514	0.60
6.United	112,673	4,395	65,530,209	0.69
7.Continental	46,771	3,051	35,215,605	0.87
8.Southwest	87,486	7,928	72,462,123	1.09
9.Delta	163,846	9,222	83,386,595	1.11
10.Alaska	24,921	1,657	14,132,047	1.17
Total	**803,344**	**33,642**	**467,204,981**	**0.72**

Los Métodos Clásicos de Pronósticos

- **Cualitativos** (subjetivos, de juicios, basados en cálculos y opiniones):
  - Proyección fundamental.
  - Investigación de mercados.
  - Consenso de grupo.
  - Analogía histórica.
  - Método Delphi.
- **Cuantitativos** (basados en métodos estadístico-matemáticos):
  - Análisis de las series de tiempo.
  - Proyección causal.
  - Modelos de simulación.

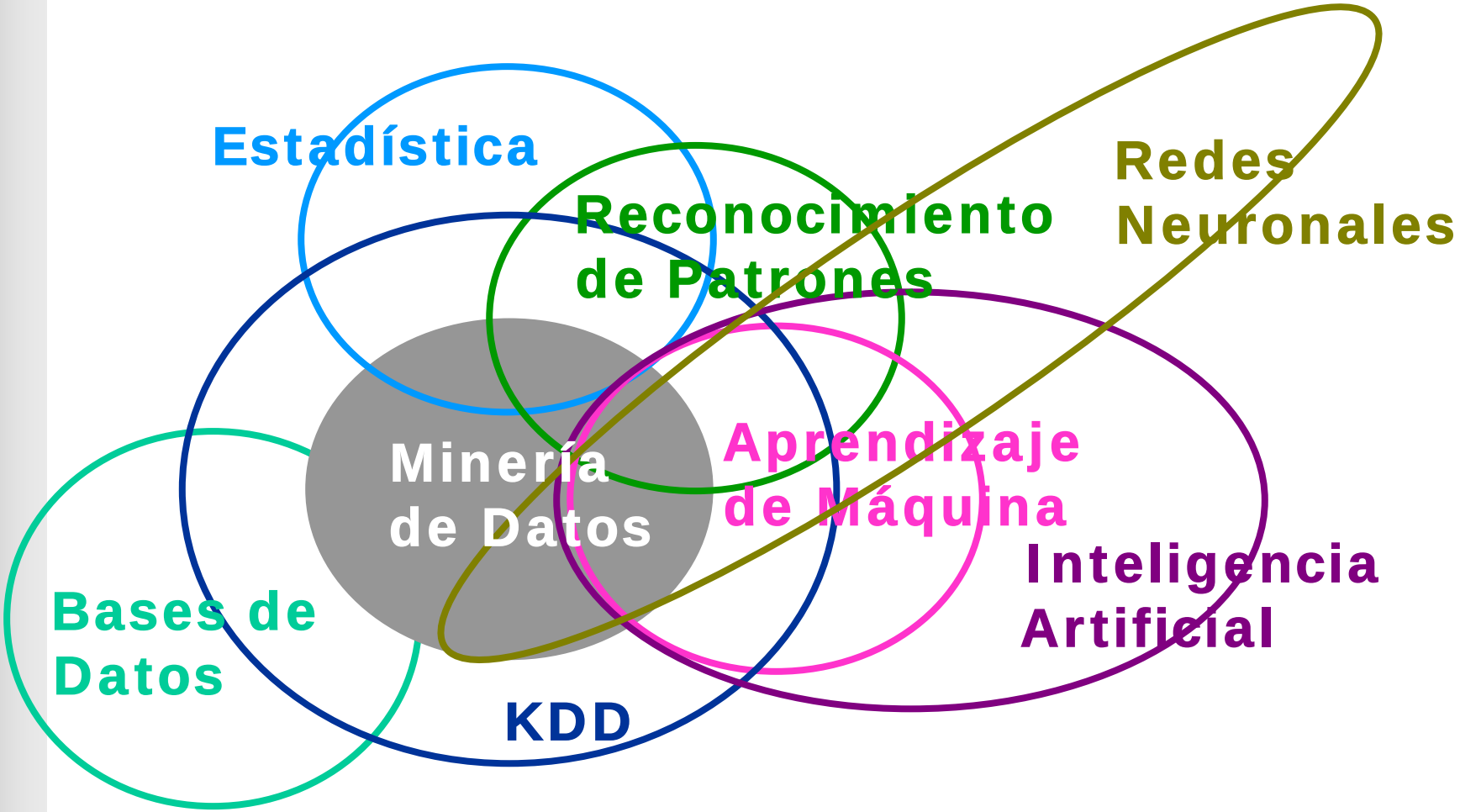
Las limitaciones de los modelos actuales de soporte

- Dirigidos a explicar el comportamiento de la variable objetivo.
- No toman en consideración las variables independientes o muy pocas de éstas.
- Los modelos matemáticos descriptivos:
  - Modelos de Markov.
  - Modelos de series de tiempo.
  - Modelos de líneas de espera, etc.
- Los modelos de investigación de mercados:
  - Síntomas, no el patrón real de comportamiento del mercado.
  - Modelos cuantitativos.
  - Modelos cualitativos.

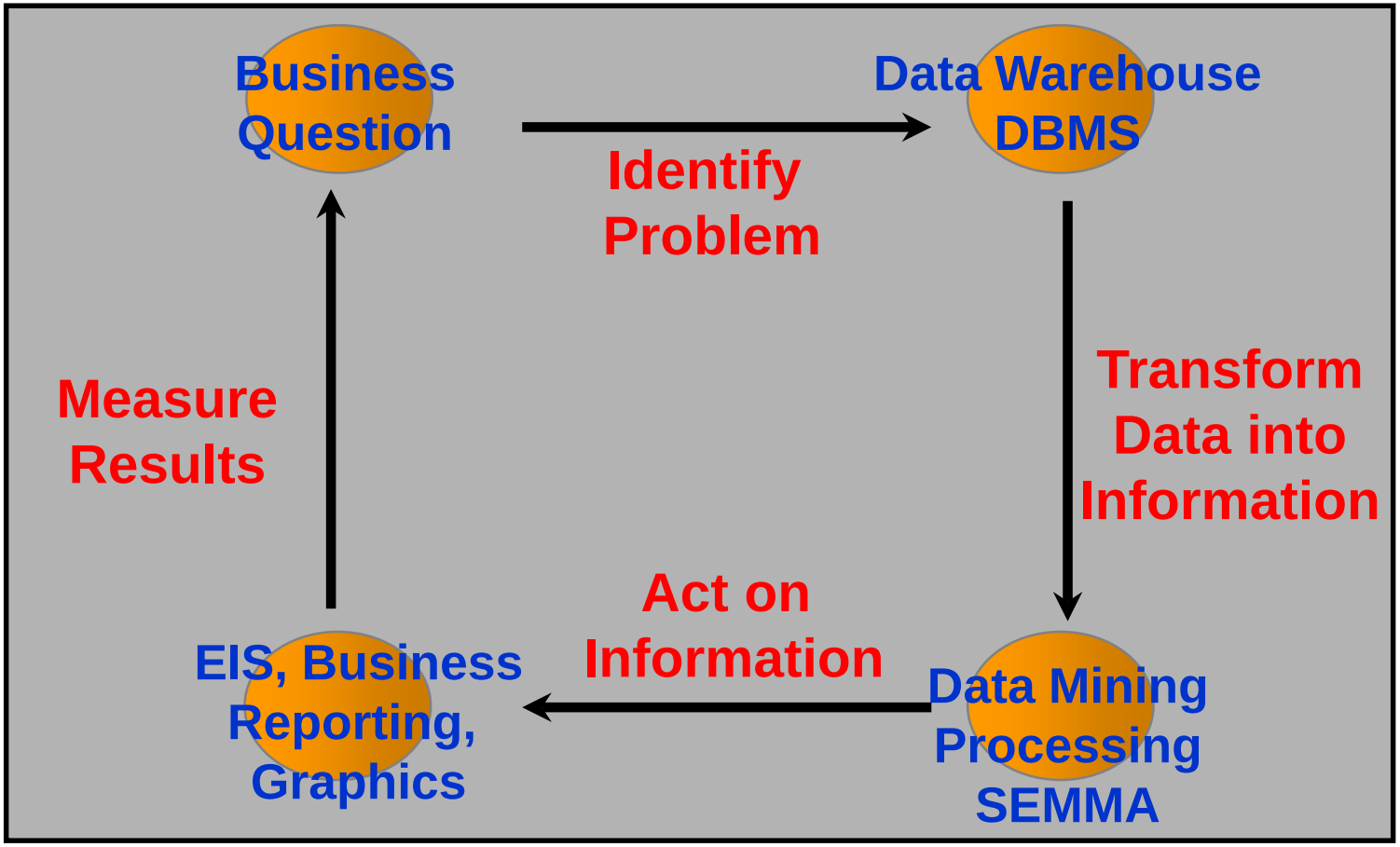
La Minería de Datos

La minería de datos es el proceso iterativo (cíclico) de selección, exploración y modelación de grandes volúmenes de datos, con el fin de revelar patrones de comportamiento antes desconocidos, y así obtener una ventaja competitiva ***sostenible*** para la organización.

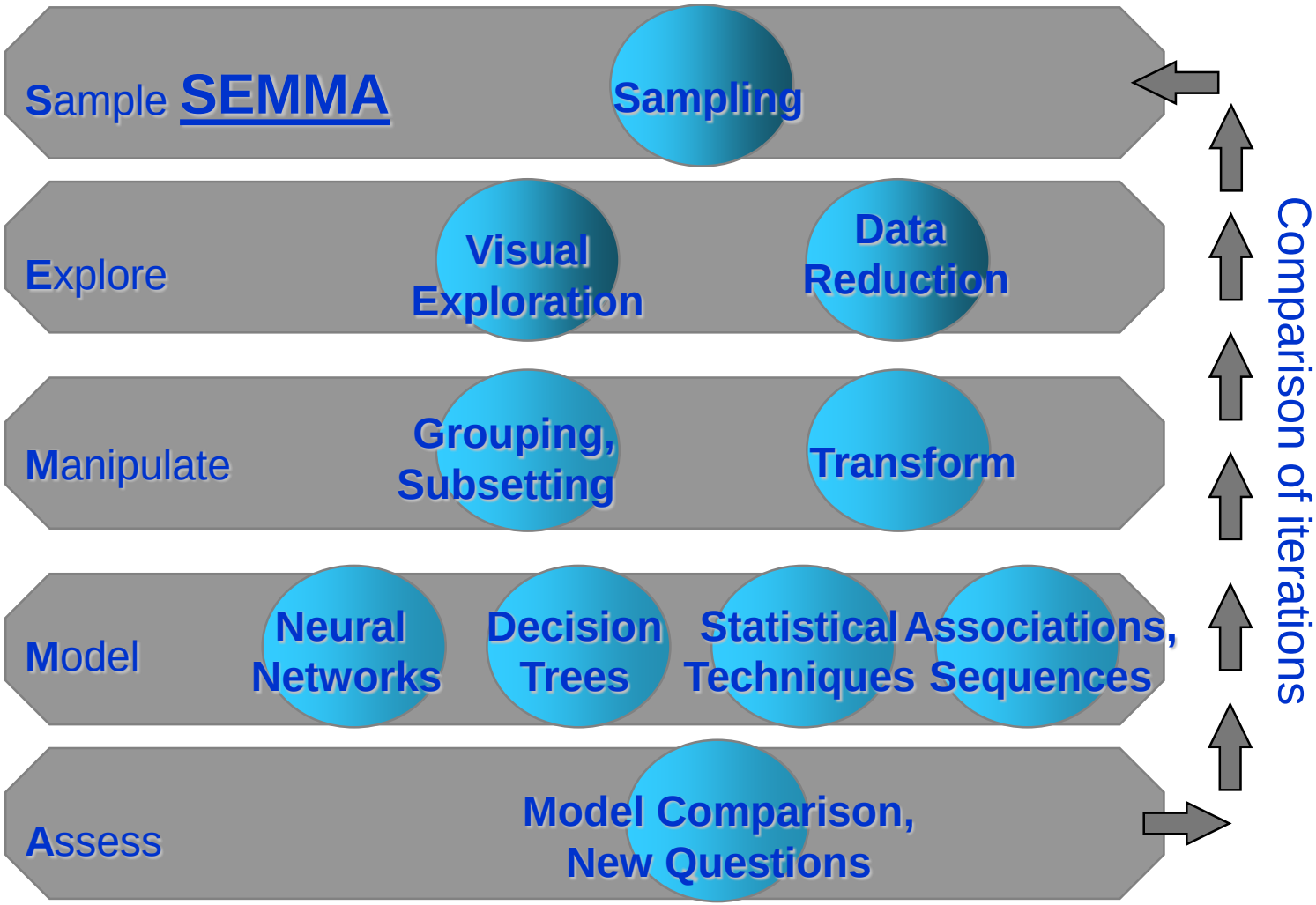
Enfoque Multidisciplinario



El Proceso de Minería de Datos



La Metodología de Investigación



Características fundamentales de los modelos de minería de datos

- **Estáticos.**
  - Estabilización de variables de comportamiento dinámico.
- **Preponderantemente estocásticos**.
  - Variables aleatorias.
- **Variables independientes.**
  - Discretas (cualitativas o no métricas).
    - – Nominales o clasificatorias
    - – Ordinales o de rangos.
  - Continuas (cuantitativas o métricas).
    - – Intervalos.
    - – Proporciones.
- **Variables dependientes** (objetivos o de respuestas).
  - Discretas (nominales (binarias) y ordinales).
  - Continuas (métricas).
- **Alta precisión y confiabilidad de las predicciones.**

Ventajas de poder desarrollar modelos predictivos in-house

- Ganancias a partir de economías de escala (muchos modelos para muchos segmentos)
- Consolidar una base de datos flexible y reutilizable (consistencia en la interpretación de los resultados de los modelos y los reportes y en la propia metodología de modelación)
- Verificar la precisión y analizar las fortalezas y debilidades de los modelos
- Reducir el acceso de extraños a información estratégica y retener las ventajas competitivas con la creación de las mejores prácticas de la compañía

Formulación del Problema de Negocio para AeroMéxico

Para determinadas condiciones del vuelo, determinar con gran exactitud y confiabilidad, cuáles son los pasajeros con mayor propensión a no presentarse, con el objetivo de estimar la sobreventa correcta para compensar las no presentaciones, de forma que se aumente el ingreso y se minimicen los costos y los inconvenientes a los pasajeros que se les niega el abordaje (denied boarding).

The Flow Diagram of Analytical Data

- Booking data warehouse o data mart.
- ETL (extraction, transformation and load) of data.
- Statistical exploration of analytical data.
- Mathematical modeling (discovery of passenger no show pattern).
- Score passengers (assignment probability for no show).
- Set the correct overbooking (probability break point)

The Booking Data Warehouse Structure (variables available from PNR)

- Booking time prior to departure (count).
- Booking class (categorical).
- Service class (categorical).
- Number of passengers in PNR (count).
- Origin city (categorical).
- Destination airport (categorical).
- Board point airport (categorical).
- Off point airport (categorical).
- Weekday of departure (categorical).

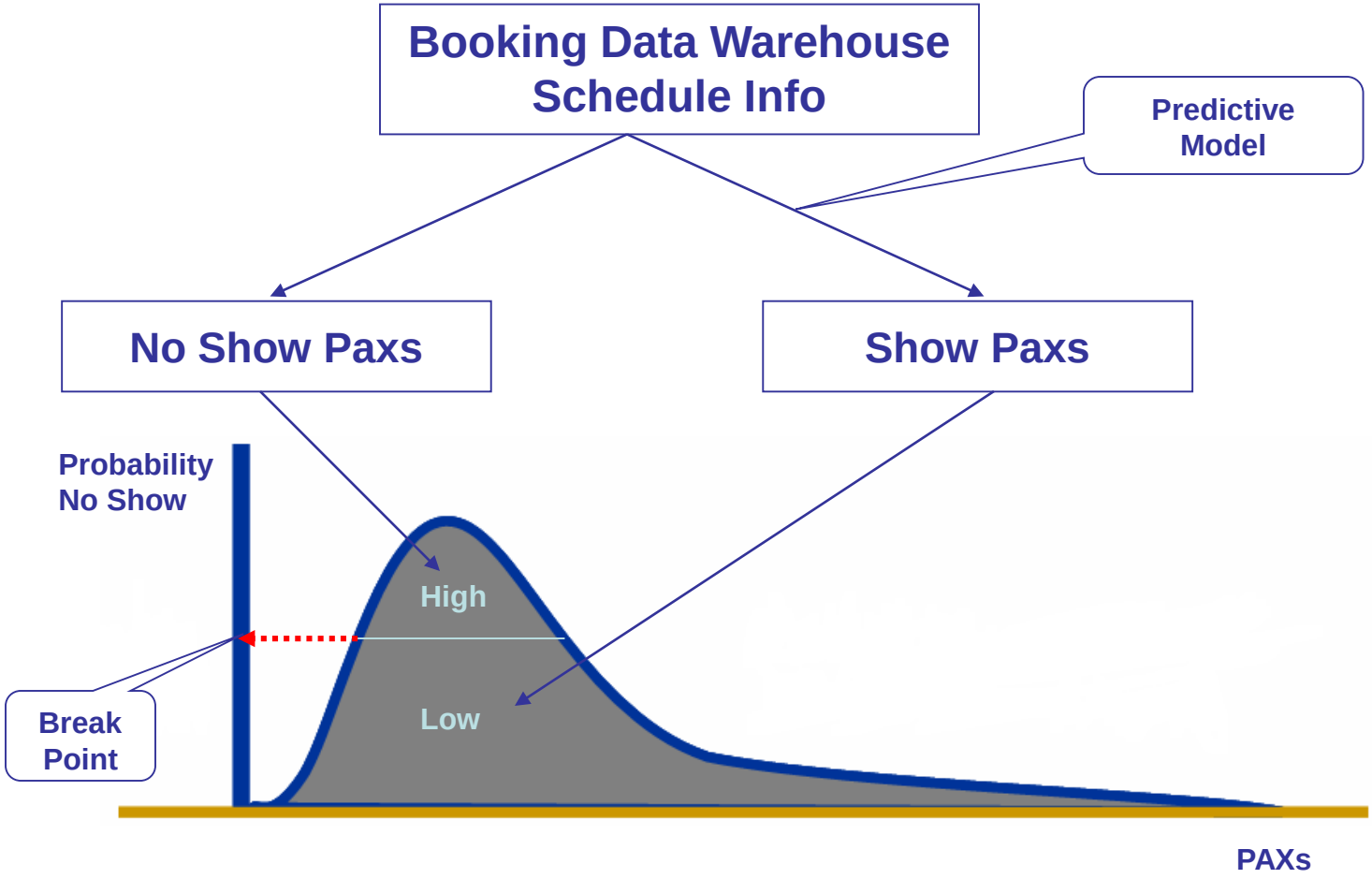
Generated Variables (use of Trip Analyzer and OD Builder algorithm)

- Was PNR split during booking history? (binary).
- Origin region (categorical).
- Destination region (categorical).
- Board point region (categorical).
- Off point region (categorical).
- Grouped Board point airports (categorical).
- Position of segment within OD (count).
- Total travel time for OD (continuous).
- Total flight time for OD (continuous).
- Flight time (segment) / Flight time (OD) (categorical).
- Connection time between segments in OD (categorical).
- More than one airline used in OD (binary).
- Is segment part of round trip? (binary).
- Does segment belong to return portion of trip? (binary).
- Purpose of trip (business, leisure, or mixed) (categorical).
- Total time for trip (continuous).
- Number of segment in trip (count).
- Number of scheduled flights per week (count).

Métodos Analíticos y Herramientas de Modelación Matemática

- La predicción probabilística de no presentación (no show passengers):
  - Árboles de decisión.
  - Regresión logística.
  - Redes neuronales.
- La metodología de modelación – SEMMA
- La herramienta de minería – SAS Enterprise Miner

Representación esquemática del modelo predictivo de “No Show”



Fragmento de la tabla de resultados del modelo predictivo

Nodo	P(No Show)	P(Show)	Pred(No Show)	Real (No Show)
37	1.0000	0.0000	1	1
37	1.0000	0.0000	1	1
37	1.0000	0.0000	1	1
21	0.8712	0.1288	1	1
21	0.8712	0.1288	1	1
21	0.8712	0.1288	1	1
19	0.8222	0.1778	1	1
19	0.8222	0.1778	1	1
19	0.8222	0.1778	1	1
30	0.6667	0.3333	1	1
30	0.6667	0.3333	1	1
30	0.6667	0.3333	1	1
36	0.4559	0.5441	0	1
36	0.4559	0.5441	0	1
36	0.4559	0.5441	0	0
20	0.2222	0.7778	0	0
20	0.2222	0.7778	0	0
20	0.2222	0.7778	0	0
31	0.1983	0.8017	0	1
31	0.1983	0.8017	0	0
31	0.1983	0.8017	0	0
4	0.0879	0.9121	0	1
4	0.0879	0.9121	0	0
4	0.0879	0.9121	0	0
6	0.0524	0.9476	0	0
6	0.0524	0.9476	0	0
6	0.0524	0.9476	0	0

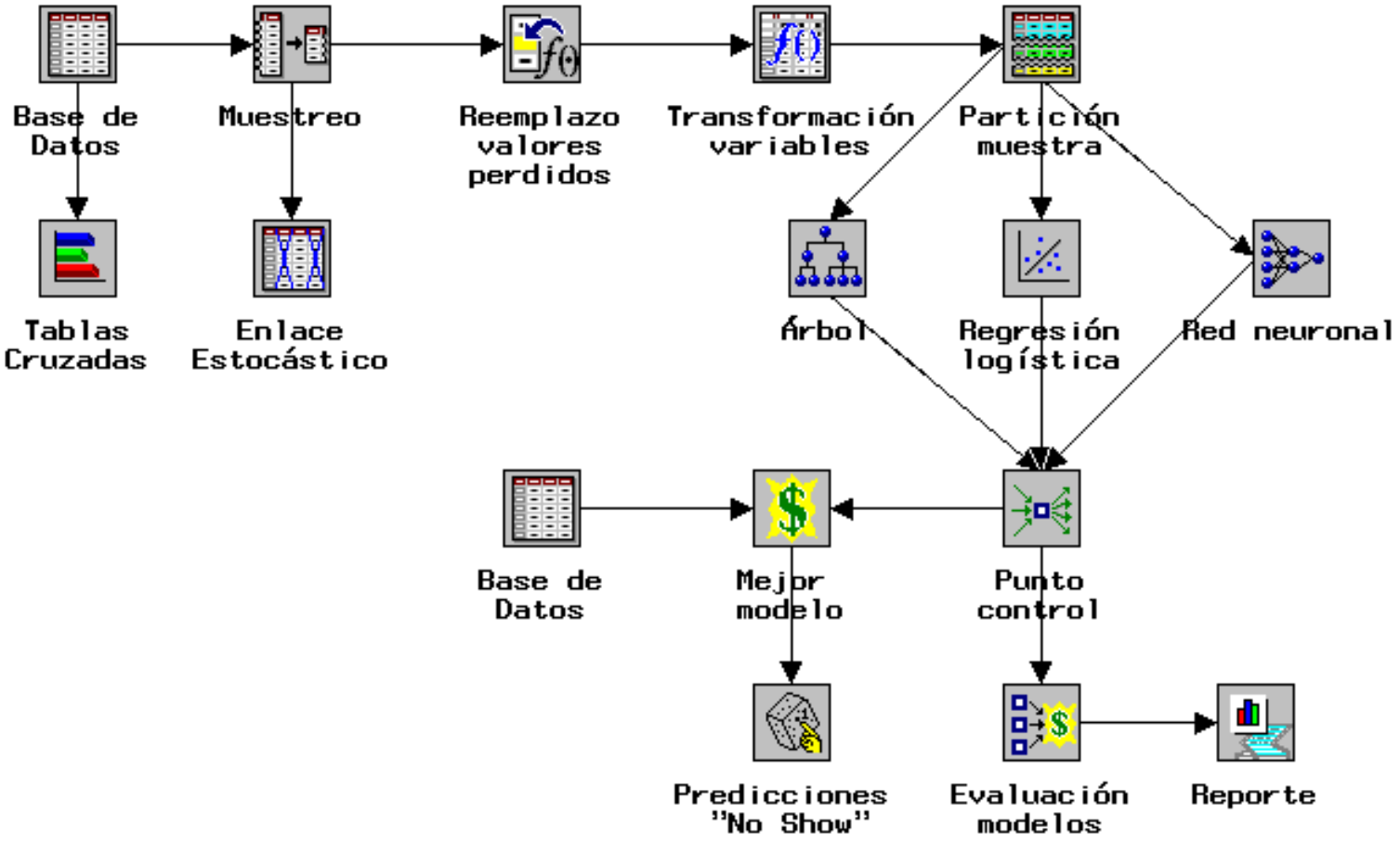
Other variables (Predictors)

Break Point

Las etapas del proceso analítico en la construcción del modelo

- Análisis Estadístico Preliminar
  - Comprensión de variables y revisión de escalas
  - Tamizado de variables
  - Construcción y análisis de la variable objetivo
- Modelación Matemática con el Minero de Datos
  - Fortaleza del enlace estocástico con la variable objetivo
  - Desarrollo de un modelo topológico
  - Desarrollo de modelos de regresión y redes
  - Comparación de modelos y elección del mejor
  - Asignación de puntajes a los pasajeros.
  - Determinación del punto de corte.
  - Estimación de la sobreventa correcta.

Esquema de construcción del modelo con el minero de SAS



Principales Resultados del Modelo

- Determinación probabilística de la propensión de lo pasajeros a no presentarse (No Show).
- Determinación de la sobreventa correcta (Overbooking).
- Mejor utilización de las capacidades de los vuelos (Spoilage).
- Disminución de los costos por rechazos a abordar (denied boarding).
- Reducción de las molestias a los pasajeros por problemas de sobreventas y rechazos.
- Aumento de la rentabilidad de los vuelos.

Conclusiones

- Los modelos predictivos de minería de datos sustituyen a los métodos de pronósticos tradicionales, ya que permiten enfocar con mayor precisión el cálculo de los pasajeros que no se presentarán (no show paxs) en los vuelos.
- El enfoque de minería de datos ya se está usando para la resolución del problema de la sobreventa correcta en cada vuelo en compañías aéreas internacionales como Continental Airlines.
- En AeroMéxico existe la posibilidad de aplicar estás técnicas de modelación predictiva de alta precisión y confiabilidad con el objetivo de aumentar la rentabilidad de los vuelos.

Datos Personales

- Viterbo H. Berberena González.
- Doctor en Ciencias Técnicas.
- Director de Minería de Datos de Pearson S.A. de C.V.
- Homero 223, PH, Colonia Polanco.
- CP 11560, México D.F.
- Teléfonos:(52)(55) 5531-5324, 5531-5560 ext. 145
- Fax: (52)(55) 5203-8230
- mailto: vberberena@pearson-research.com
- USA phone: (305) 390 8242 Miami.
- 1-800 711 7709 –TOLL FREE.