XII
CONGRESSO NAZIONALE degli ATTUARI

Data Science applicata alle scienze attuariali

Applicazioni nell'underwriting, nel ramo danni e nel ramo vita
Metodi di regressione vs. Tecniche di machine learning

Marco Aleandri

22 novembre 2018

Il progetto di ricerca

- Cosa?
  - Concetti di base (partizione del dataset, misure di performance, software, ecc.)
  - Unsupervised learning (regole di associazione, PCA, cluster analysis, ecc.)
  - Supervised learning (GLM, k-NN, CART, reti neurali, ensembles, ecc.)
- Come?
  - Applicazione nell'underwriting: *customer management*
  - Applicazione nel danni: *individual claim reserving*
  - Applicazione nel vita: *dynamic policyholder behaviour*
- Perché?
  - Adattabilità ai dati (?)
  - Spiegazione di informazione (?)
  - Interpretabilità dei risultati (?)

Il progetto di ricerca

- Cosa?
  - Concetti di base (partizione del dataset, misure di performance, software, ecc.)
  - Unsupervised learning (regole di associazione, PCA, cluster analysis, ecc.)
  - Supervised learning (GLM, k-NN, CART, reti neurali, ensembles, ecc.)
- Come?
  - Applicazione nell'underwriting: *customer management*
  - Applicazione nel danni: *individual claim reserving*
  - Applicazione nel vita: *dynamic policyholder behaviour*
- Perché?
  - Adattabilità ai dati (?)
  - Spiegazione di informazione (?)
  - Interpretabilità dei risultati (?)

Customer Management – Market Basket Analysis

id	*antecedent*		*consequent*	*support*	*confidence*	*lift*	*count*
1	VehAge_fac=[0,5]	=>	Fire=Y	0,33	0,98	1,40	7047
2	VehClass=Expensive	=>	Fire=Y	0,01	0,85	1,22	197
3	VehClass=Medium	=>	Fire=Y	0,03	0,84	1,20	677
4	VehClass=Medium high	=>	Fire=Y	0,02	0,84	1,20	478
5	VehPower_fac=(3,6]	=>	Fire=Y	0,19	0,83	1,19	3988
6	VehClass=Medium low	=>	Fire=Y	0,09	0,82	1,18	1843
7	VehGas=Diesel	=>	Fire=Y	0,32	0,82	1,17	6814
8	VehPower_fac=(6,9]	=>	Fire=Y	0,01	0,79	1,13	260
9	DrivAge_fac=(60,70]	=>	Fire=Y	0,04	0,78	1,12	812
10	Channel=B	=>	Fire=Y	0,10	0,78	1,12	2037

id	*antecedent*		*consequent*	*support*	*confidence*	*lift*	*count*
1	VehAge_fac=[0,5]	=>	DamAll=Y	0,30	0,89	1,82	6437
2	DrivAge_fac=(70,100]	=>	DamAll=Y	0,01	0,66	1,34	232
3	VehClass=Expensive	=>	DamAll=Y	0,01	0,66	1,34	152
4	VehClass=Medium	=>	DamAll=Y	0,02	0,65	1,33	526
5	VehClass=Medium low	=>	DamAll=Y	0,07	0,64	1,31	1446
6	VehPower_fac=(3,6]	=>	DamAll=Y	0,14	0,64	1,31	3074
7	VehGas=Diesel	=>	DamAll=Y	0,25	0,63	1,28	5242
8	VehClass=Medium high	=>	DamAll=Y	0,02	0,63	1,28	358
9	DrivAge_fac=(60,70]	=>	DamAll=Y	0,03	0,63	1,28	651
10	BonusMalus_fac=(0,50]	=>	DamAll=Y	0,23	0,60	1,22	4849

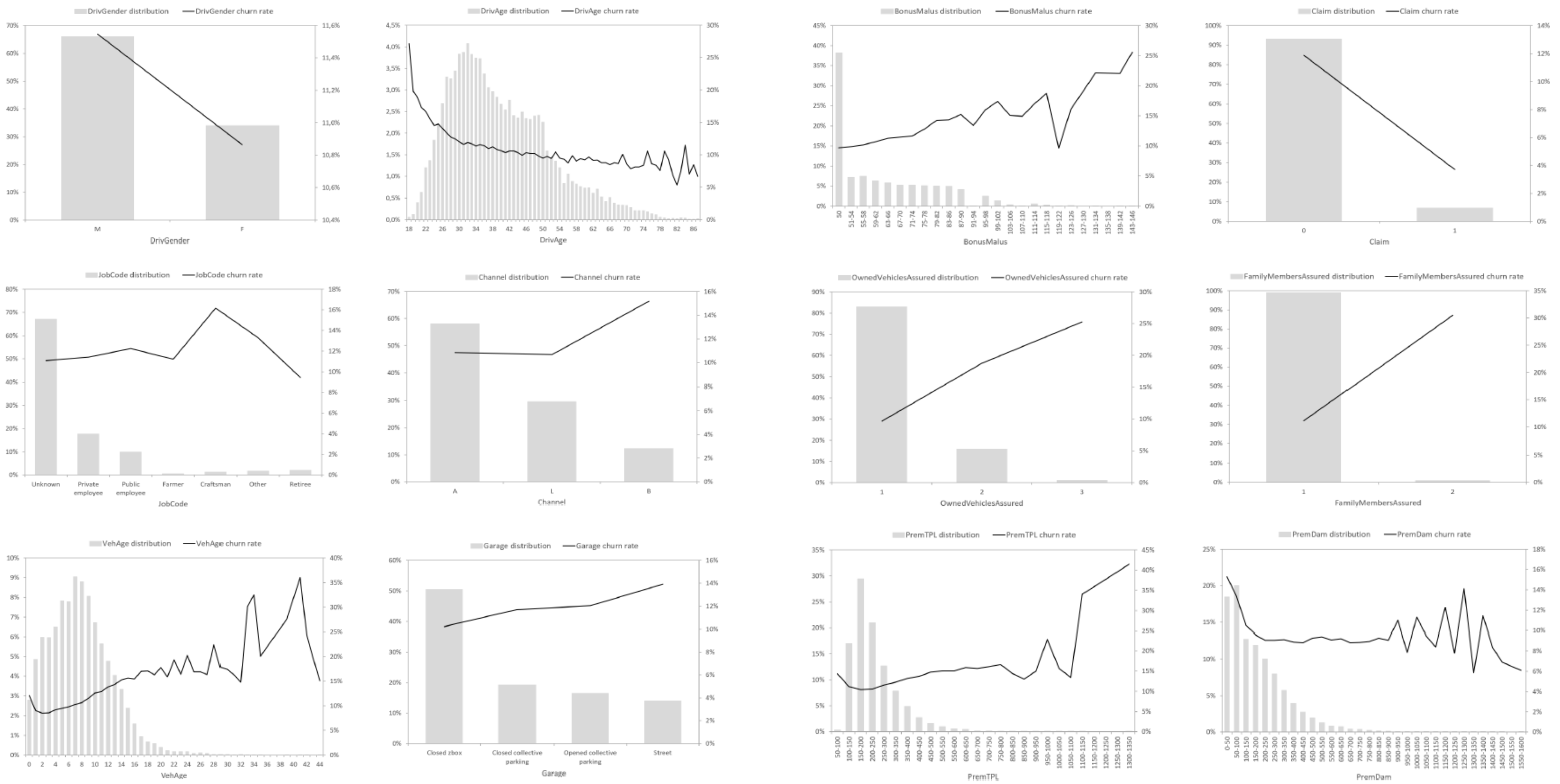
id	*antecedent*		*consequent*	*support*	*confidence*	*lift*	*count*
1	VehAge_fac=[0,5]	=>	Theft=Y	0,33	0,98	1,39	7059
2	VehClass=Expensive	=>	Theft=Y	0,01	0,84	1,19	194
3	VehClass=Medium high	=>	Theft=Y	0,02	0,83	1,18	475
4	VehClass=Medium	=>	Theft=Y	0,03	0,83	1,17	668
5	VehPower_fac=(3,6]	=>	Theft=Y	0,19	0,82	1,17	3960
6	DrivAge_fac=(70,100]	=>	Theft=Y	0,01	0,82	1,16	289
7	VehClass=Medium low	=>	Theft=Y	0,09	0,82	1,16	1835
8	VehGas=Diesel	=>	Theft=Y	0,32	0,81	1,16	6798
9	DrivAge_fac=(60,70]	=>	Theft=Y	0,04	0,81	1,15	843
10	VehAge_fac=(5,10]	=>	Theft=Y	0,32	0,79	1,12	6786

id	*antecedent*		*consequent*	*support*	*confidence*	*lift*	*count*
1	Acc2=N	=>	Acc1=Y	0,38	0,54	1,43	8048
2	OwnedVehiclesAssured=3	=>	Acc1=Y	0,01	0,46	1,21	119
3	JobCode=Other	=>	Acc1=Y	0,01	0,44	1,16	155
4	Region=Paris area	=>	Acc1=Y	0,06	0,43	1,13	1359
5	Area=A8	=>	Acc1=Y	0,02	0,42	1,11	338
6	JobCode=Public employee	=>	Acc1=Y	0,04	0,42	1,10	875
7	OwnedVehiclesAssured=2	=>	Acc1=Y	0,07	0,42	1,10	1396
8	MaritalStatus=Married	=>	Acc1=Y	0,03	0,42	1,10	604
9	MaritalStatus=Cohabiting	=>	Acc1=Y	0,08	0,41	1,09	1808
10	JobCode=Private employee	=>	Acc1=Y	0,07	0,41	1,08	1523

id	*antecedent*		*consequent*	*support*	*confidence*	*lift*	*count*
1	VehAge_fac=[0,5]	=>	Windscreen=Y	0,34	0,99	1,16	7131
2	VehClass=Expensive	=>	Windscreen=Y	0,01	0,94	1,10	217
3	DrivAge_fac=(70,100]	=>	Windscreen=Y	0,02	0,93	1,09	330
4	VehPower_fac=(3,6]	=>	Windscreen=Y	0,21	0,93	1,09	4465
5	VehClass=Medium	=>	Windscreen=Y	0,04	0,93	1,09	749
6	DrivAge_fac=(60,70]	=>	Windscreen=Y	0,05	0,93	1,08	962
7	VehClass=Medium high	=>	Windscreen=Y	0,02	0,92	1,08	527
8	VehClass=Medium low	=>	Windscreen=Y	0,10	0,92	1,08	2069
9	VehGas=Diesel	=>	Windscreen=Y	0,36	0,91	1,07	7597
10	BonusMalus_fac=(0,50]	=>	Windscreen=Y	0,35	0,90	1,06	7353

id	*antecedent*		*consequent*	*support*	*confidence*	*lift*	*count*
1	Region=South West	=>	Acc2=Y	0,07	0,52	1,72	1413
2	Acc1=N	=>	Acc2=Y	0,30	0,48	1,61	6397
3	Area=A7	=>	Acc2=Y	0,06	0,42	1,41	1189
4	VehUsage=Professional	=>	Acc2=Y	0,01	0,41	1,36	190
5	MaritalStatus=Unknown	=>	Acc2=Y	0,26	0,38	1,28	5499
6	JobCode=Unknown	=>	Acc2=Y	0,26	0,38	1,28	5499
7	BonusMalus_fac=(100,150]	=>	Acc2=Y	0,01	0,35	1,15	107
8	Garage=Street	=>	Acc2=Y	0,05	0,34	1,13	1012
9	LicenceNb=4	=>	Acc2=Y	0,01	0,34	1,13	164
10	Garage=Closed collective parking	=>	Acc2=Y	0,06	0,34	1,12	1368

Customer Management – Renewal Rate Prediction



Il progetto di ricerca

- Cosa?
  - Concetti di base (partizione del dataset, misure di performance, software, ecc.)
  - Unsupervised learning (regole di associazione, PCA, cluster analysis, ecc.)
  - Supervised learning (GLM, k-NN, CART, reti neurali, ensembles, ecc.)
- Come?
  - Applicazione nell’underwriting: *customer management*
  - Applicazione nel danni: *individual claim reserving*
  - Applicazione nel vita: *dynamic policyholder behaviour*
- Perché?
  - Adattabilità ai dati (?)
  - Spiegazione di informazione (?)
  - Interpretabilità dei risultati (?)

Individual Claim Reserving – Closing Delay vs. Claim Amount

$$\hat{R}_i = \sum_{k=0}^{3} \hat{p}_i(\Delta_i = k)\hat{C}_{\Delta_i=k}$$

Actual Class	Predicted Class 0	1	2	3	Cases Number	Errors Number	Errors Percentage
0	97	1.805	73	12	1.987	1.890	95,12%
1	130	3.339	329	61	3.859	520	13,47%
2	83	2.082	445	71	2.681	2.236	83,40%
3	44	1.027	207	126	1.404	1.278	91,03%
Total					**9.931**	**5.924**	**59,65%**

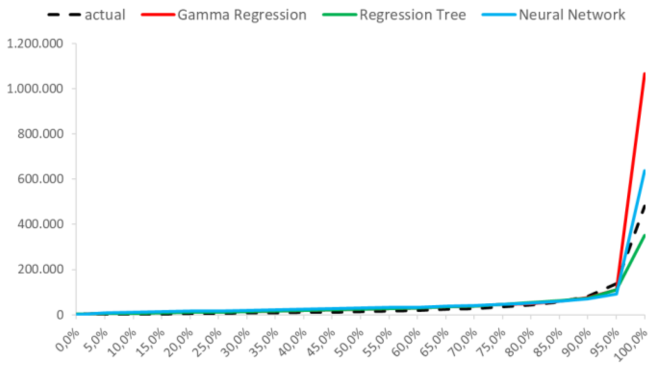
Multinomial regression summary results using training data

Actual Class	Predicted Class 0	1	2	3	Cases Number	Errors Number	Errors Percentage
0	72	1.194	44	15	1.325	1.253	94,57%
1	92	2.185	220	37	2.534	349	13,77%
2	52	1.444	287	60	1.843	1.556	84,43%
3	29	664	155	70	918	848	92,37%
Total					**6.620**	**4.006**	**60,51%**

Multinomial regression summary results using validation data

reporting year	closing delay 0	1	2	3
1993	26.038.265	77.236.474	74.285.051	58.268.484
1994	16.746.725	42.454.730	42.485.823	55.798.479
1995	6.076.308	20.958.156	41.895.020	50.580.334
1996	4.090.537	21.665.535	46.736.025	8.158.885

Actual data



reporting year	closing delay 0	1	2	3
1993	23.942.277	72.096.426	85.372.988	68.135.627
1994	15.615.349	43.607.867	49.123.608	69.977.076
1995	6.294.610	21.590.823	41.199.048	51.380.998
1996	4.338.858	23.040.826	46.178.322	10.158.513

Gamma regression

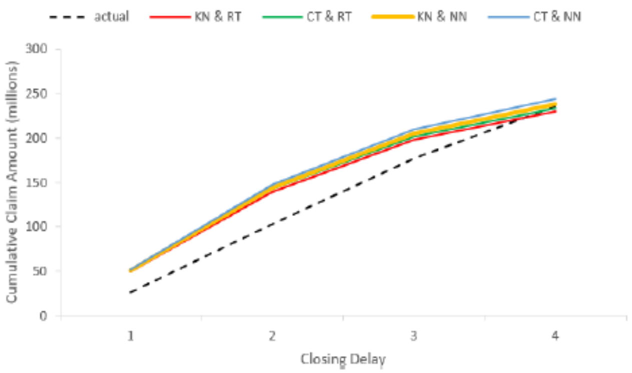
reporting year	closing delay 0	1	2	3
1993	27.777.729	74.715.426	74.860.543	53.671.653
1994	17.317.427	44.645.051	41.115.866	56.014.481
1995	7.819.221	22.443.511	40.888.916	51.092.375
1996	5.285.040	22.293.701	46.183.852	9.048.303

Regression tree

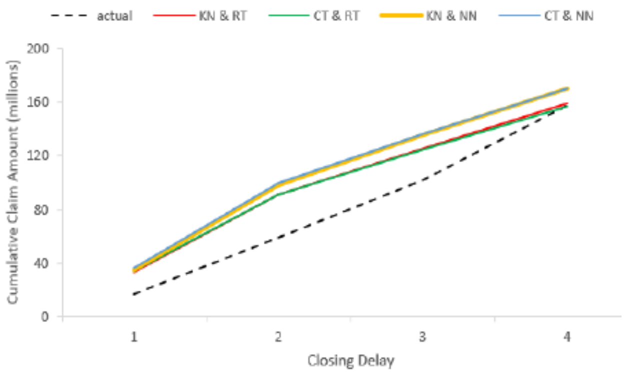
reporting year	closing delay 0	1	2	3
1993	30.323.454	80.122.598	74.090.121	56.198.161
1994	21.876.621	49.791.426	44.957.040	51.693.100
1995	8.007.124	25.605.931	47.106.288	49.418.199
1996	5.358.636	27.349.566	46.177.586	8.455.383

Neural network

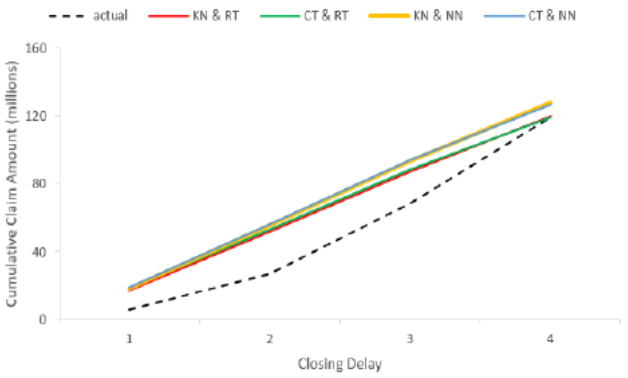
Individual Claim Reserving – Reserve Estimation



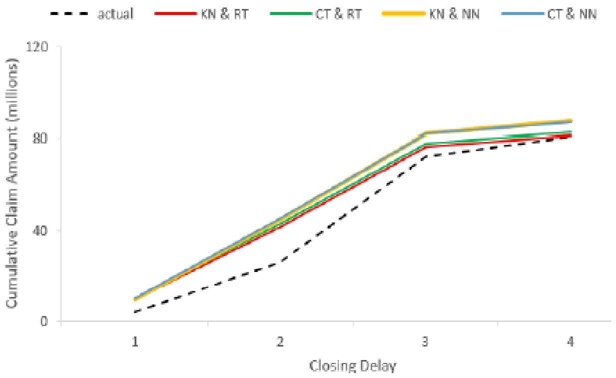
Cumulative amount in 1993



Cumulative amount in 1994



Cumulative amount in 1995



Cumulative amount in 1996

Actual claim amount 593.474.833	*gamma regression*	*regression tree*	*neural network*
multinomial regression	638.895.354	594.577.625	631.222.248
naive Bayes	631.416.644	582.831.820	626.198.292
nearest neighbours	633.263.257	590.737.696	623.292.207
classification tree	641.129.119	592.631.628	627.681.248

Δ%	*gamma regression*	*regression tree*	*neural network*
multinomial regression	7,65%	0,19%	6,36%
naive Bayes	6,39%	-1,79%	5,51%
nearest neighbours	6,70%	-0,46%	5,02%
classification tree	8,03%	-0,14%	5,76%

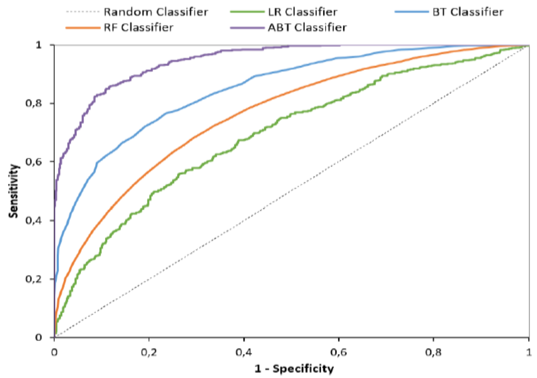
Overall predictive performance

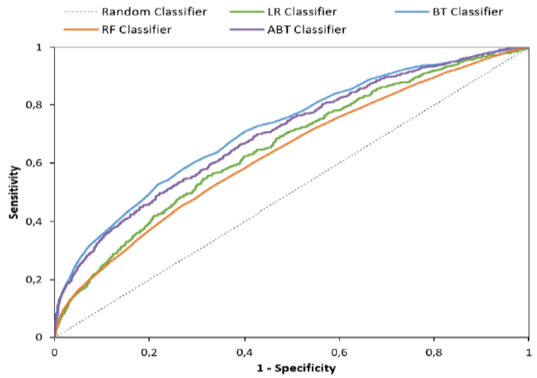
Il progetto di ricerca

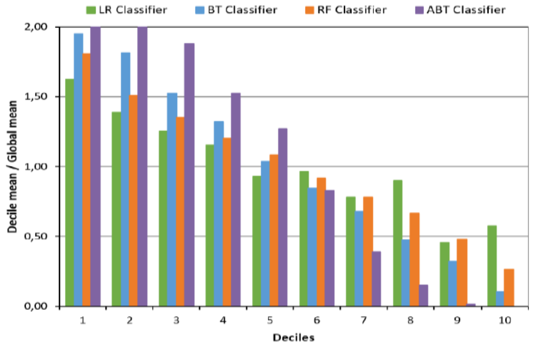
- Cosa?
  - Concetti di base (partizione del dataset, misure di performance, software, ecc.)
  - Unsupervised learning (regole di associazione, PCA, cluster analysis, ecc.)
  - Supervised learning (GLM, k-NN, CART, reti neurali, ensembles, ecc.)
- Come?
  - Applicazione nell'underwriting: *customer management*
  - Applicazione nel danni: *individual claim reserving*
  - Applicazione nel vita: *dynamic policyholder behaviour*
- Perché?
  - Adattabilità ai dati (?)
  - Spiegazione di informazione (?)
  - Interpretabilità dei risultati (?)

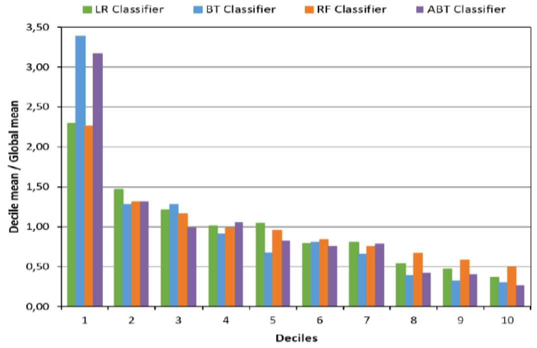
Dynamic Policyholder Behaviour – Lapse Prediction

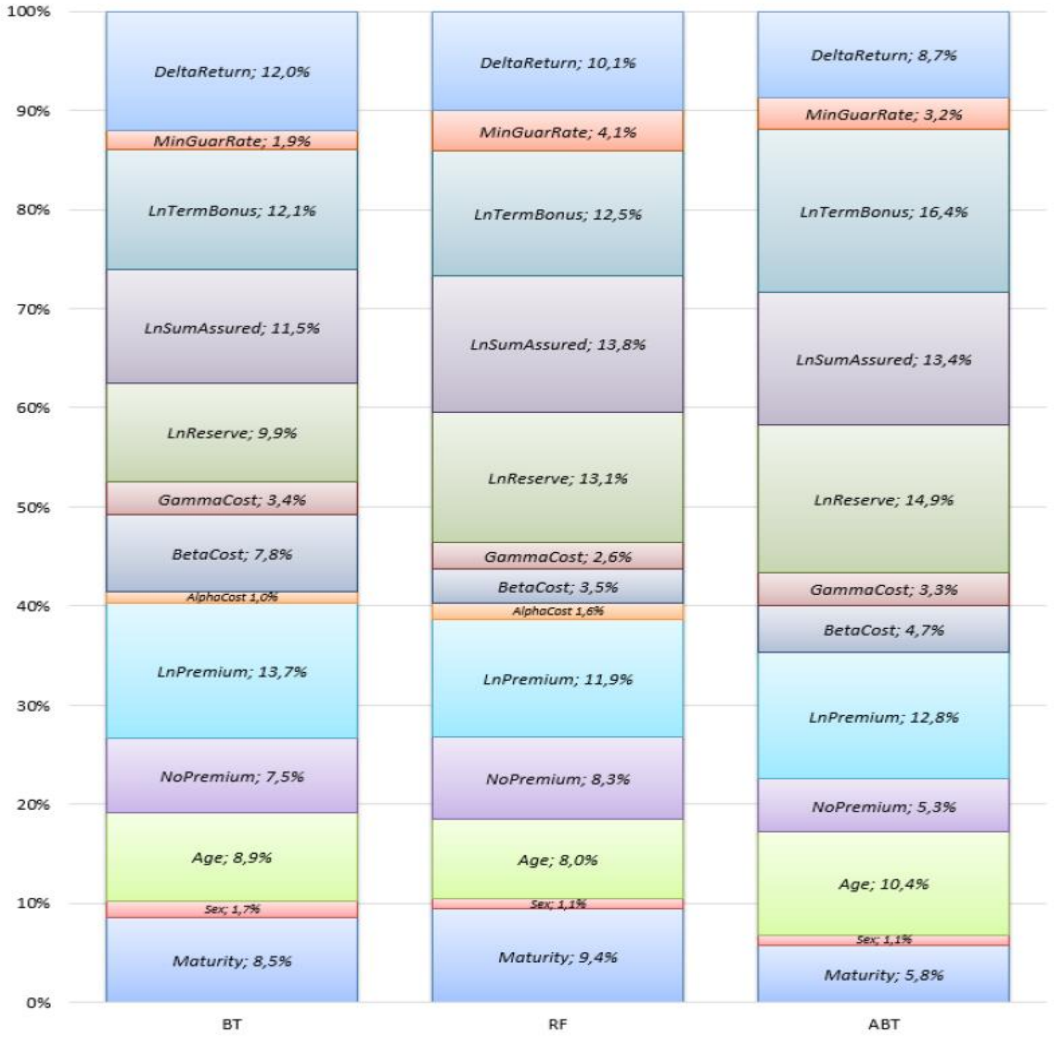
Method	Logistic Regression	**Bagging Trees**	Random Forest	Boosting Trees
AUC_T	69%	**80%**	77%	92%
AUC_V	65%	**72%**	64%	70%



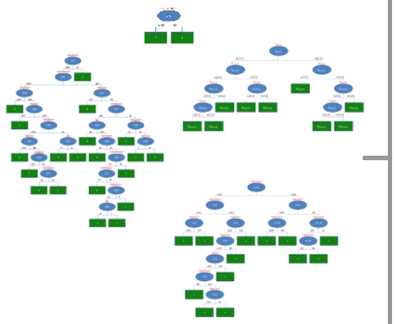








Dynamic Policyholder Behaviour – Performance Metrics



$$p_t(ce) := \frac{1}{m}\sum_{i=1}^{m} p_t(i;ce)$$

$$D[ce,p(ce)] := \sum_{t=1}^{n} p_t(ce)D_t(ce)$$

$$D[k,p(ce)] := \sum_{t=1}^{n} p_t(ce)D_t(k) \longrightarrow \frac{1}{N}\sum_{k=1}^{N} D[k,p(ce)] \longrightarrow D[ce,p(ce)] - D[N,p(ce)]$$

$$p_t(k) := \frac{1}{m}\sum_{i=1}^{m} p_t(i;k)$$

$$D[k,p(k)] := \sum_{t=1}^{n} p_t(k)D_t(k) \longrightarrow \frac{1}{N}\sum_{k=1}^{N} D[k,p(k)] \longrightarrow D[ce,p(ce)] - D[N,p(N)]$$

$$1 = \underbrace{\frac{TVOG[N,p(ce)]}{TVOG[N,p(N)]}}_{\text{guarantee}} + \underbrace{\frac{V_{PHB}(p)}{TVOG[N,p(N)]}}_{\text{PHB}}$$

$$D[k,t^*(k)] := D_{t^*(k)}(k) \longrightarrow \frac{1}{N}\sum_{k=1}^{N} D[k,t^*(k)] \longrightarrow D[ce,p(ce)] - D[N,t^*(N)]$$

$$1 = \overbrace{\frac{TVOG[N,p(ce)]}{TVOG[N,t^*(N)]}}^{\text{guarantee}} + \overbrace{\frac{V_{PHB}(t^*)}{TVOG[N,t^*(N)]}}^{\text{PHB}}$$

Dynamic Policyholder Behaviour – Profit vs. TVOG

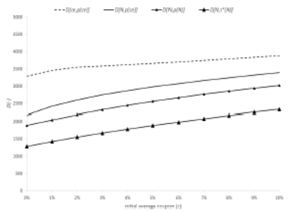


Fig. 4.40: Profit metrics by initial average coupon (FTSE MIB case)

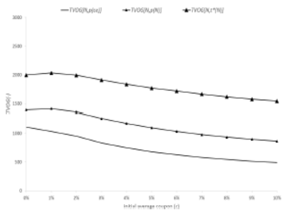


Fig. 4.41: TVOG metrics by initial average coupon (FTSE MIB case)

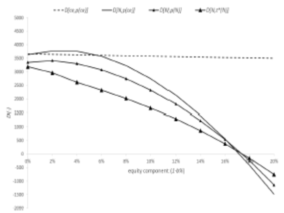


Fig. 4.46: Profit metrics by equity percentage (FTSE MIB case)

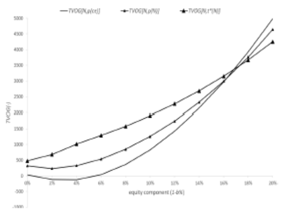


Fig. 4.47: TVOG metrics by equity percentage (FTSE MIB case)

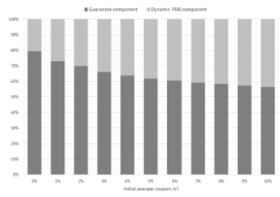


Fig. 4.68: $TVOG[N, p(N)]$ guarantee-PHB split by initial average coupon (FTSE MIB case)

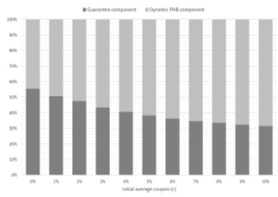


Fig. 4.69: $TVOG[N, t^*(N)]$ guarantee-PHB split by initial average coupon (FTSE MIB case)

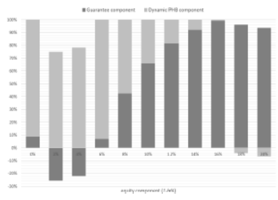


Fig. 4.74: $TVOG[N, p(N)]$ guarantee-PHB split by equity percentage (FTSE MIB case)

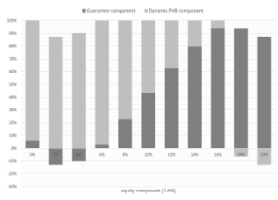


Fig. 4.75: $TVOG[N, t^*(N)]$ guarantee-PHB split by equity percentage (FTSE MIB case)

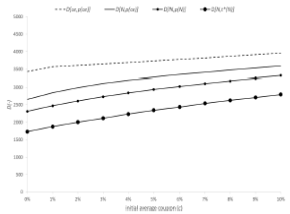


Fig. 4.42: Profit metrics by initial average coupon (EURO STOXX 50 case)

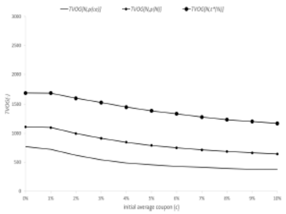


Fig. 4.43: TVOG metrics by initial average coupon (EURO STOXX 50 case)

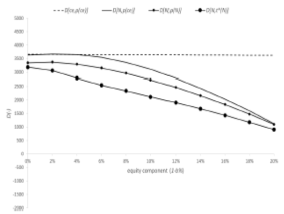


Fig. 4.48: Profit metrics by equity percentage (EURO STOXX 50 case)

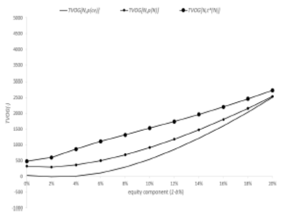


Fig. 4.49: TVOG metrics by equity percentage (EURO STOXX 50 case)

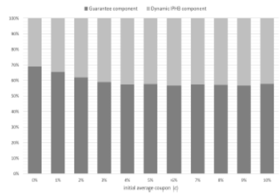


Fig. 4.70: $TVOG[N, p(N)]$ guarantee-PHB split by initial average coupon (EURO STOXX 50 case)

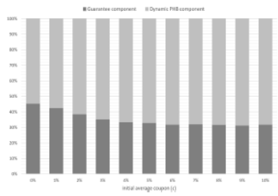


Fig. 4.71: $TVOG[N, t^*(N)]$ guarantee-PHB split by initial average coupon (EURO STOXX 50 case)

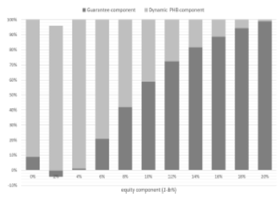


Fig. 4.76: $TVOG[N, p(N)]$ guarantee-PHB split by equity percentage (EURO STOXX 50 case)

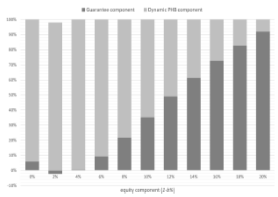


Fig. 4.77: $TVOG[N, t^*(N)]$ guarantee-PHB split by equity percentage (EURO STOXX 50 case)

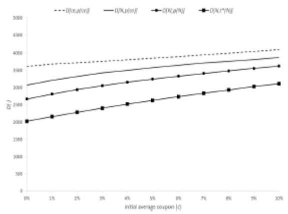


Fig. 4.44: Profit metrics by initial average coupon (S&P 500 case)

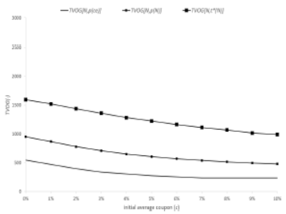


Fig. 4.45: TVOG metrics by initial average coupon (S&P 500 case)

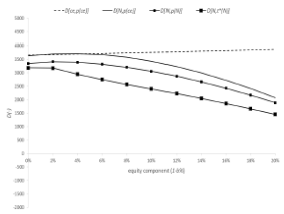


Fig. 4.50: Profit metrics by equity percentage (S&P 500 case)

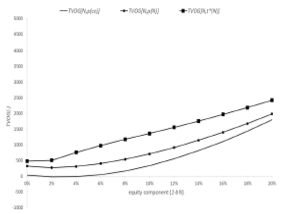


Fig. 4.51: TVOG metrics by equity percentage (S&P 500 case)

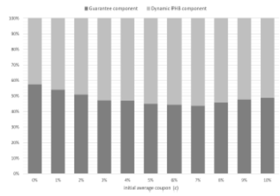


Fig. 4.72: $TVOG[N, p(N)]$ guarantee-PHB split by initial average coupon (S&P 500 case)

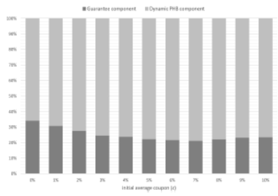


Fig. 4.73: $TVOG[N, t^*(N)]$ guarantee-PHB split by initial average coupon (S&P 500 case)

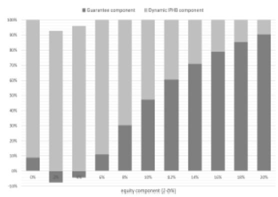


Fig. 4.78: $TVOG[N, p(N)]$ guarantee-PHB split by equity percentage (S&P 500 case)

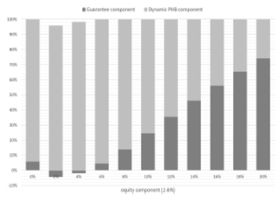
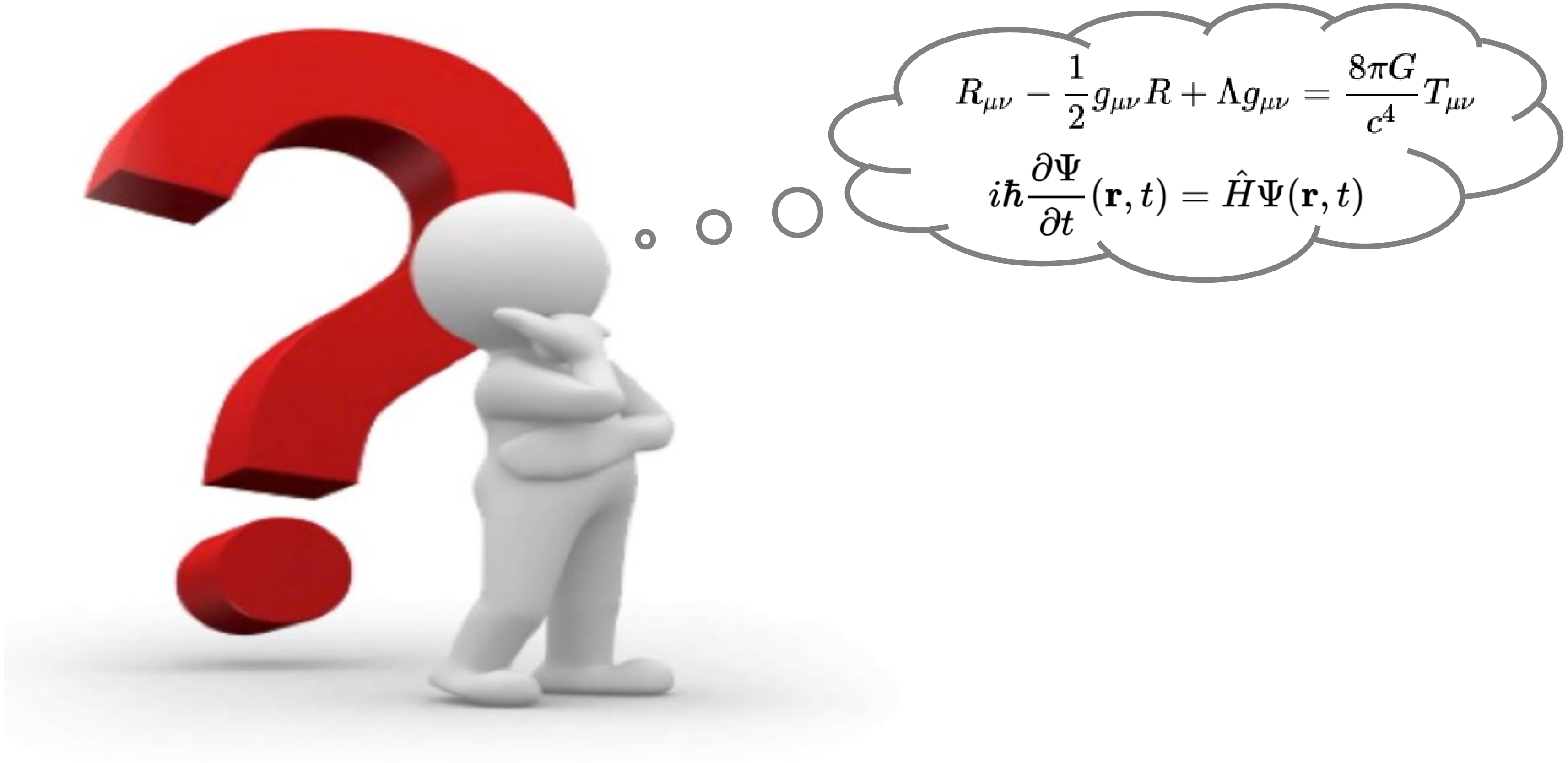


Fig. 4.79: $TVOG[N, t^*(N)]$ guarantee-PHB split by equity percentage (S&P 500 case)

Q&A



$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$
$$i\hbar \frac{\partial \Psi}{\partial t}(\mathbf{r}, t) = \hat{H} \Psi(\mathbf{r}, t)$$