

# Prédiction de la réussite des étudiants à partir des patterns de connexion dans un espace numérique de travail

Yvonnick Noël, Univ. Rennes 2 - LP3C

Festival d'Automne du Numérique - Paris, 19 novembre 2018

# Sommaire

- 1 Enjeux
- 2 Situation étudiée
- 3 Analyses
  - Analyse du décrochage
  - Analyse de la réussite à l'examen
- 4 Conclusions



## Diversité des publics dans le supérieur

- Constat : on trouve à l'université des publics **très hétérogènes en niveau** :
  - bacheliers maîtrisant parfaitement les outils scolaires,
  - étudiants aux acquis plus fragiles,
  - personnes en reprise d'études, aux niveaux très hétérogènes,
  - étudiants étrangers.
- Pour **améliorer la réussite** des étudiants et prévenir le décrochage, il faut :
  - développer des outils de détection du **décrochage**,
  - développer des modules de positionnement et de formation **adaptatifs**.

- Cours de statistiques en **L3 de psychologie**, en 2016-2017.
- Les ressources sont sur une page Moodle et comprennent :
  - une **rubrique générale** avec : objectifs de la formation et forum des nouvelles,
  - spécifiquement, **pour chaque année** de la L1 au M2 :
    - une présentation des objectifs,
    - une page de ressources (polys, annales, diaporamas CM, fiches TD),
    - un forum d'entraide,
    - des tests automatisés.

## Modèles binomiaux et gaussiens (semestre 5)

Yvonnick Noël, Camilo Charron, Alessandro Guida et Serban Musca

-  Objectifs du cours [STATS-L3]
-  Polycopié, diapos de cours, annales d'examen, fiches TD, programmes pour calculettes [STATS-L3]
-  Forum d'entraide des étudiants de L3
-  Test n°1/6 (Licence 3 semestre 5)  
**Accès restreint** Non disponible à moins que : Vous soyez membre d'un groupe de **L3 Psycho S5**
-  Test n°2/6 (licence 3 semestre 5)  
**Accès restreint** Non disponible à moins que : Vous soyez membre d'un groupe de **L3 Psycho S5**
-  Test n°3/6 (licence 3 semestre 5)  
**Accès restreint** Non disponible à moins que : Vous soyez membre d'un groupe de **L3 Psycho S5**
-  Test n°4/6 (licence 3 semestre 5)  
**Accès restreint** Non disponible à moins que : Vous soyez membre d'un groupe de **L3 Psycho S5**
-  Test n°5/6 (licence 3 semestre 5)  
**Accès restreint** Non disponible à moins que : Vous soyez membre d'un groupe de **L3 Psycho S5**
-  Test n°6/6 (licence 3 semestre 5)  
**Accès restreint** Non disponible à moins que : Vous soyez membre d'un groupe de **L3 Psycho S5**

## Base 1 : traces sur Moodle

DATE	JOUR	HEURE	NUMERO	ELEMENT	DUREE	SESSION
2016-09-16	15	16:04:00	20005420	Page principale		1
2016-09-16	15	16:04:38	20005420	Forum L2		1
2016-09-16	15	16:04:52	20005420	Forum L2	14	1
2016-09-17	16	08:14:05	20005420	Page principale		2
2016-10-08	37	07:00:24	20005420	Forum L2		3
2016-10-08	37	07:00:46	20005420	Forum L2	22	3
2016-10-08	37	07:00:47	20005420	Forum L2	1	3
2016-11-23	83	16:05:15	20005420	Forum L2		4
2016-11-23	83	16:05:50	20005420	Forum L2	35	4
2016-11-23	83	16:13:12	20005420	Ressources L2	442	4
2016-11-23	83	16:20:08	20005420	Ressources L2	416	4
2016-10-28	57	09:22:40	20200449	Page principale		1
2016-10-28	57	09:22:56	20200449	Ressources L3 S5	16	1
2016-11-08	68	12:29:12	20200449	Page principale		2
2016-11-08	68	12:29:48	20200449	Page principale	36	2
2016-11-08	68	12:30:26	20200449	Page principale	38	2
2016-11-08	68	12:30:39	20200449	Ressources L3 S5	13	2
2016-11-08	68	17:35:37	20200449	Page principale	18298	3
2016-11-08	68	17:35:44	20200449	Ressources L3 S5	7	3
2016-11-08	68	17:36:06	20200449	Ressources L3 S5	22	3
2016-12-14	104	10:43:42	20200449	Page principale		4
2016-12-14	104	10:44:10	20200449	Ressources L3 S5	28	4
2016-12-14	104	10:44:30	20200449	Page principale	20	4
2016-12-14	104	10:44:32	20200449	Sujets de TER	2	4

## Base 2 : résultats Apogée

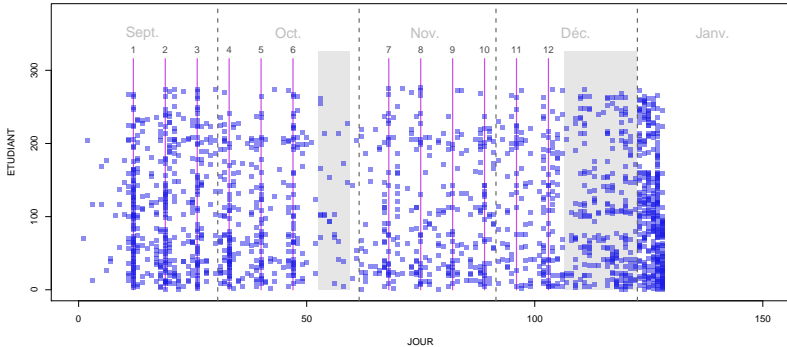
NUMERO	STATUT	NOTE STATS	NOTE SEM.	ABSENTSTAT
20206477	SUP	4	6	NON
21606526	ASSIDU	1.5	8.336	NON
21405305	ASSIDU	5	11.211	NON
21402871	ASSIDU	3	8.254	NON
21208410	ASSIDU	0	5.042	OUI
21608607	SUP	0	0	OUI
21612221	SUP	0	6.012	OUI
21604897	SUP	2	9.61	NON
21305793	ASSIDU	0	9.948	NON
20603776	SUP	0	1.95	OUI
21402000	ASSIDU	10.5	12.631	NON
97004410	SUP	0	3.033	OUI
21412639	ASSIDU	0	6.219	NON
21613009	SUP	0	2.6	OUI
21406075	ASSIDU	11.5	14.513	NON
21407844	ASSIDU	1.5	11.258	NON
21516191	SUP	2.5	7.888	NON
21408215	ASSIDU	7.5	11.342	NON
21612162	SUP	2.5	8.713	NON
21613448	SUP	0	0	OUI
21500862	SUP	3.5	13.008	NON
21402581	ASSIDU	7.5	11.9	NON
21414158	SUP	2.5	10.658	NON
21405349	ASSIDU	6	11.888	NON
21612257	ASSIDU	1.25	11.67	NON
20305124	SUP	1	9.44	NON
21413933	SUP	0	0	OUI

## Méthodologie

- On exploite deux bases : la base des **traces individuelles** sur Moodle et la base Apogée des **résultats à l'examen**, dans la matière et pour l'ensemble du semestre.
- La clé de croisement des deux bases est le **numéro anonyme** de l'étudiant.
- La méthodologie utilisée pour la prévision du décrochage (présence/absence à l'examen) et de la réussite (note) en fin de semestre, à partir des patterns de connexion journaliers, est celle des **modèles linéaires généralisés pénalisés** (Friedman, Hastie & Tibshirani, 2008) à travers le package `glmnet` pour R.



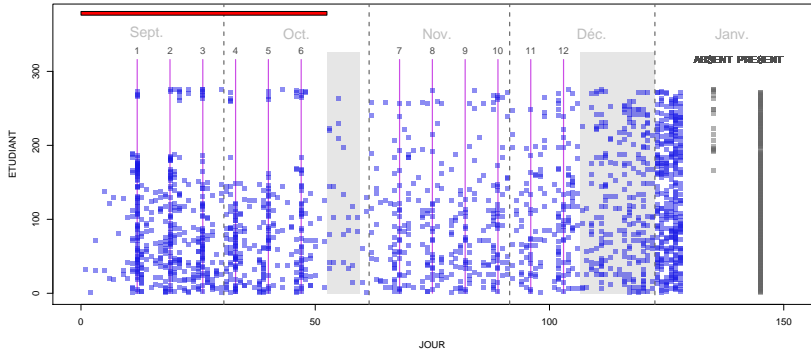
# Diagramme de connexions



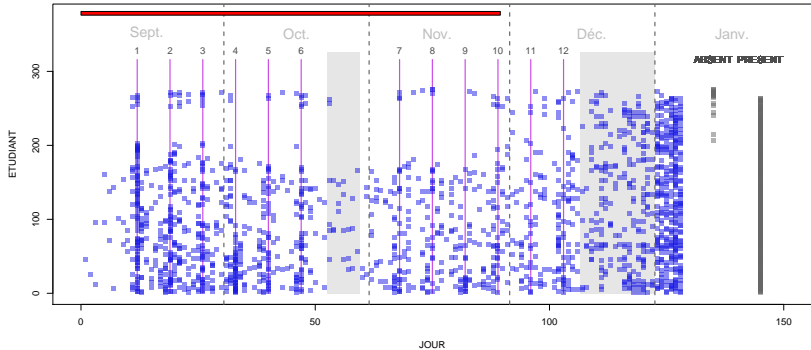
## Objectifs

- Prévoir qui sera **absent de l'épreuve finale** du 7 janvier, à partir du **seul profil de fréquence de connexions**. sans regarder la nature des ressources cliquées.
- Nous souhaitons construire un **outil diagnostique** permettant d'alerter les étudiants **assez tôt** dans le semestre.
- On essaie tour à tour ( $N = 276$  étudiants assidus) :
  - une prévision à partir des connexions **jusqu'aux vacances d'octobre** (abandons nombreux après),
  - une prévision à partir des connexions **jusqu'à fin novembre** (un mois pour se ressaisir),
  - une prévision à partir des connexions **jusqu'à avant les vacances de Noël** (trois semaines pour se ressaisir),
  - une prévision à partir des connexions **sur tout le semestre**.

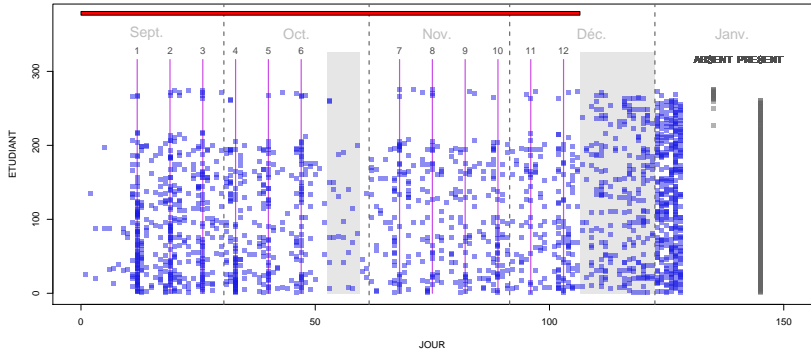
# Prévision sur la période 1



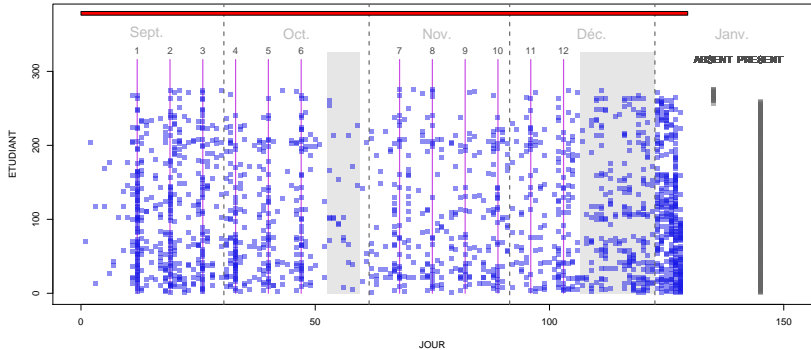
## Prévision sur la période 2



## Prévision sur la période 3



# Prévision sur la période 4



## Synthèse

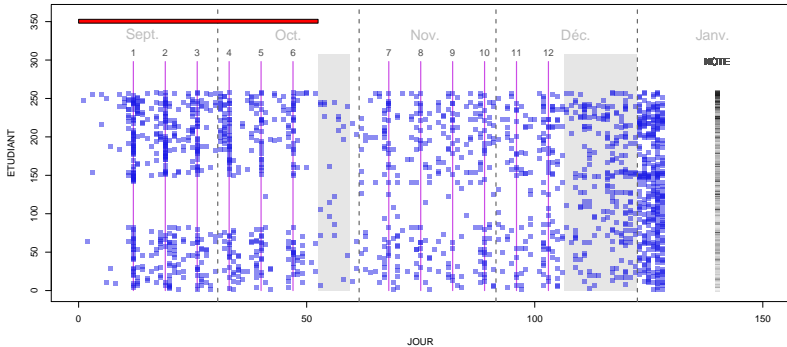
Période	% expliq.	% Bien classés	Faux positifs	Faux négatifs
<21 octobre	35%	77.7%	0%	77.7%
<30 novembre	64%	97.8%	0%	33.3%
<17 décembre	83.4%	98.9%	0%	16.6%
<7 janvier	95.3%	99.6%	0%	5.5%

## Conclusions

- L'analyse sur la totalité des données est **étonnamment bonne** (presque 100% de diagnostics corrects), compte tenu de la nature très superficielle des données (**fréquence d'activité** uniquement).
- Davantage de données sont requises pour permettre néanmoins une **prévision plus précoce** dans le semestre.

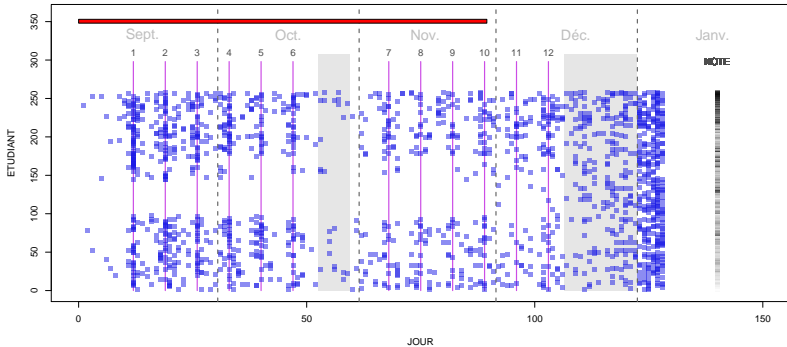


# Prévision sur la période 1



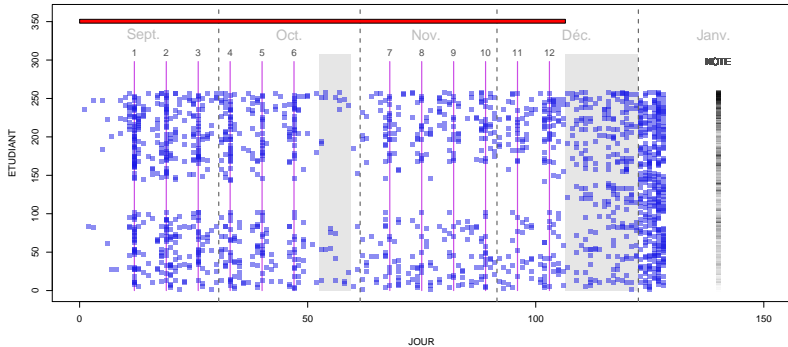
$$R^2 = 0.223$$

# Prévision sur la période 2



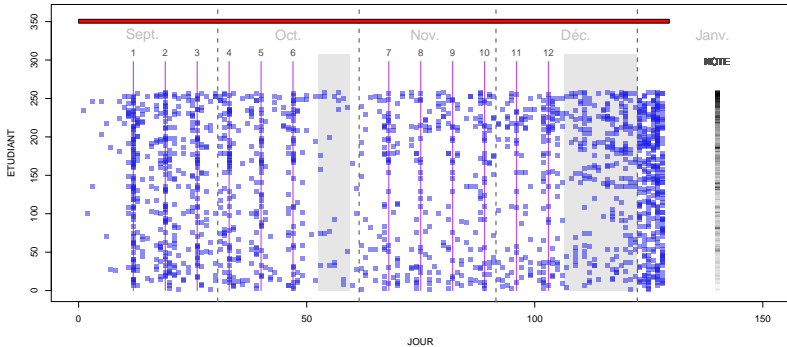
$$R^2 = 0.410$$

# Prévision sur la période 3



$$R^2 = 0.484$$

# Prévision sur la période 4



$$R^2 = 0.552$$

## Commentaires

- Les étudiants qui ont des bons résultats ont des **débits de connexion importants** (haut du graphique).
- Mais cette remarque **ne s'inverse pas** : des étudiants à haut débit de connexion n'ont pas nécessairement de bonnes notes (bas du graphique).
- Les **patterns** associés aux hautes notes montrent :
  - une synchronisation avec les jours du cours (assiduité), matérialisés par les 12 traits verticaux,
  - le travail pendant les vacances de Noël.

## Conclusions

- Ces résultats sont encourageants : avec une information purement quantitative sur les **débits de fréquentation**, on parvient à des niveaux raisonnables de prédiction sur le décrochage et sur la réussite.
- On peut penser que la prédiction sera encore meilleure en intégrant des **informations qualitatives** précises sur :
  - les ressources qui ont été cliquées,
  - les chaînes de transitions de ressource en ressource (analyses markoviennes).