

# Análisis basados en n-gramas para la clasificación del nivel de Burnout en académicos universitarios

Mauricio Castro, David Pinto, Darnes Vilariño, Mireya Tovar

Benémerita Universidad Autónoma de Puebla,  
Facultad de Ciencias de la Computación, México

mayka2000@yahoo.com.mx,

{davideduardopinto,dvilarinoayala,mireyatovar}@gmail.com

<http://www.lke.buap.mx/>

**Resumen.** En este artículo se presenta un estudio del nivel de estrés sobre la base de los comentarios abiertos emitidos por académicos universitarios usando reconocimiento del lenguaje natural. El enfoque presenta la aplicación de un algoritmo de clasificación automática y el análisis de características para dicho clasificador basadas en n-gramas de palabras. Los resultados de dos experimentos son mostrados: 1) el impacto en el incremento en el tamaño del  $n$ -grama en el proceso de clasificación, y 2) la verificación en la separación de dos clases, una con bajo y la otra con alto nivel de Burnout. Los resultados obtenidos muestran un porcentaje de clasificación correcta cercano al 60% cuando se consideran tres clases: alto, medio y bajo nivel de Burnout, mientras que dicho porcentaje se incrementa a 79% cuando únicamente se consideran dos clases: alto y bajo nivel de Burnout.

**Palabras clave:** Clasificación automática, nivel de Burnout, N-gramas de palabras.

## N-gram based Analysis for Classifying the Burnout Level of University Professors

**Abstract.** In this paper we present a study of stress level using a set of open comments written by university professors employing natural language processing techniques. The approach uses a classification algorithm and feature analysis based on word n-grams. We show the results obtained in two experiments: 1) The impact of increasing the n-gram size in the classification process, and 2) An analysis of the degree of separation of two classes: one with low and another with high level of Burnout. The obtained results show a classification accuracy close to 60% when using three classes (high, medium and low level of Burnout), whereas, that percentage increases to 79% when we employed only two classes (high and low level of Burnout).

**Keywords.** Machine learning, Burnout level, word N-grams.

## 1. Introducción

El término estrés fue inicialmente acuñado por Hans Selye en 1936 al experimentar con ratas expuestas a diferentes condiciones externas nocivas para ellas. Selye le llamó “Stress” a la reacción de las ratas a esas condiciones externas, y describió la conducta de las ratas en su trabajo [6].

En su momento, el estudio del estrés se reducía a establecer los cambios fisiológicos que sufría el ser humano expuesto a amenazas del medio. Sin embargo, desde finales de la década de los 60, el estrés se asocia a amenazas de carácter organizacional. Freudenberg [1] estudia la conducta de personas con altos niveles de estrés y que fueron expuestas a duros retos laborales, llamándole “Síndrome de Burnout” a este padecimiento. Freudenberg describe este padecimiento como sentimientos de cansancio emocional y frustración asociados a una carga irracional de trabajo.

En el año de 1976, Cristina Maslash define al Síndrome de Burnout como cansancio Emocional, despersonalización y baja realización profesional, y de esta forma se abren las perspectivas de estudios cuantitativos que culminan para el año 1981 con la creación de un instrumento que está compuesto de 22 ítems organizados de la siguiente forma: 9 ítems relacionados con la dimensión de cansancio emocional, 5 con la dimensión de despersonalización y 8 con la dimensión de realización profesional. A este instrumento se le ha llamado “Maslash Burnout Inventory” [2,3] (en adelante MBI) y fue validado por medio del análisis factorial con una muestra de 1,025 personas de diferentes ámbitos laborales pero todos con profesiones relacionadas con el área de servicios como son: enfermeras, maestros, trabajadores sociales etc. subrayando las autoras, fuertes correlaciones de los resultados del análisis factorial con el contexto demográfico. Un análisis de las principales fortalezas y debilidades del instrumento se pueden ver en [4].

Paralelamente a los estudios cuantitativos de este síndrome con los instrumentos descritos anteriormente, en la literatura se habla de validaciones externas, es decir de observadores externos que por medio de estudios clínicos determinan si una persona tiene o no el síndrome. Así, en el mismo trabajo [2], se realiza una “validación externa” con ayuda de 40 médicos de la salud mental, los cuales realizaron una evaluación clínica de los encuestados con el instrumento MBI.

La identificación automática del síndrome de Burnout en las personas puede ser de gran beneficio pues permitiría atender este padecimiento y evitar complicaciones a largo plazo en su desarrollo personal y profesional. En este trabajo de investigación, existe el interés de detectar el nivel de Burnout en académicos universitarios utilizando los comentarios de texto escritos por los mismos docentes en un cuestionario sobre su percepción del desempeño y del ambiente laboral.

El resto de este trabajo se encuentra dividido como sigue. En la Sección 2. se presenta la metodología propuesta para el desarrollo de esta investigación. La Sección 3. muestra los datos utilizados, los experimentos, los resultados obtenidos, así como una discusión de dichos resultados. Finalmente, en la Sección 4. se dan las conclusiones de este trabajo y los experimentos planteados a futuro.

## 2. Metodología

En diciembre del 2012 se aplicó el instrumento “Maslash Burnout Inventory” a 2,943 profesores de diferentes facultades de la Benemérita Universidad Autónoma de Puebla (BUAP), con la intención de buscar las causas del bajo rendimiento que se había observado años atrás en algunos docentes, y las quejas de estudiantes de mal trato de algunos profesores.

Al aplicársele a este instrumento el análisis factorial exploratorio y observar una nueva configuración de las dimensiones que resultaban validadas se generó un nuevo instrumento con 21 elementos, ordenando las dimensiones de la siguiente forma:

- 5 Elementos de la dimensión de cansancio emocional
- 8 Elementos de la dimensión despersonalización
- 8 Elementos de la dimensión realización profesional

Se calculó el Alfa de Gronbach a cada dimensión, variando ésta de entre 0.702 y 0.860. Con este nuevo instrumento se hizo el estudio de 2,943 docentes de la BUAP de un total de 5,395 trabajadores académicos. Los docentes pertenecen a 31 unidades académicas de entre las cuales, 8 unidades son del nivel medio superior (20.3% de la población), el resto pertenecen al nivel superior; de estos el 45.8% son del género femenino. Las edades oscilan entre los 23 y los 80 años con una antigüedad promedio en la BUAP de 12 años y 7.8 fuera de la BUAP. El 80% dice poseer nivel superior a licenciatura, y de estos el 13.7% tienen título de doctor con un 7.9% en proceso de titulación. El 67% tiene hijos y el 17.8% tiene más de 2 hijos. En el año 2012, el promedio de los grupos atendidos es de 32 y de alumnos atendidos 172 (el 52% de la planta docente atiende grupos de más de 30).

**Tabla 1.** Los resultados obtenidos.

Nivel de Burnout	Número de profesores	Porcentaje
<b>Libre</b>	789	26.8%
<b>Bajo</b>	865	29.4%
<b>Medio</b>	892	30.3%
<b>Alto</b>	350	11.9%
<b>Grave</b>	47	1.6%
<b>Total</b>	2,943	

Como puede verse en la Tabla 1, un 13.5% de docentes manifiestan altos niveles del síndrome de Burnout. Al final del test que se aplicó a los 2,943 docentes se les pidió que voluntariamente dieran un comentario abierto acerca de ellos o de lo que les pareciera negativo o positivo de su ambiente laboral, obteniendo en total 1,944 comentarios abiertos.

Así, el problema planteado en este trabajo de investigación sería:

- Puede detectarse el síndrome de burnout a partir de estos comentarios?, o

- Puede determinarse la no existencia del síndrome a partir de la forma en que se expresan en los comentarios abiertos?

Para este propósito, es necesario establecer una metodología para la ejecución de los experimentos que lleven a solucionar o acercarse a una solución para el problema planteado. Así, se plantea iniciar los experimentos estableciendo un conjunto de datos que emplee solamente dos clases: alto y bajo nivel de Burnout. A partir de dichos datos deseamos verificar si es posible detectar el nivel de Burnout de un académico universitario. Para tal objeto usaremos una técnica basada en aprendizaje automático, la cual considera utilizar parte de los datos como conjunto de entrenamiento y el resto como conjunto de prueba.

La Tabla 1 presenta las cinco clases que fueron usadas en el proceso de evaluación del nivel de Burnout en académicos universitarios mediante un análisis basado en un cuestionario. Son clases discretas que no son del todo excluyentes dado que, por ejemplo, un profesor con nivel grave de Burnout podría considerarse con un nivel alto y por tanto, podríamos unir a estas dos clases en un solo nivel. Así, para los experimentos planteados en este trabajo de investigación se decidió agrupar las cinco clases mostradas en la Tabla 1 en solamente tres clases: “Bajo”, “Medio” y “Alto”. Estas clases se muestran en la Tabla 2 y son el resultado de agrupar las clases “Libre” y “Bajo” en una sola clase llamada “Bajo”, y agrupar las clases “Alto” y “Grave” en una sola clase llamada “Alto”; la clase “Medio” se mantiene tal cual. La columna 2 de la Tabla 2 muestra la nueva distribución de las clases, así como la cantidad de profesores asociados con cada clase.

Es importante destacar que de los 2,943 profesores que fueron evaluados con el instrumento, solamente 2,035 de ellos escribieron un comentario abierto. Aun así, un análisis desarrollado sobre estos datos mostró que en varios casos, el mismo profesor había evaluado más de una vez (datos repetidos), por lo que se procedió a dejar solamente aquellos profesores que pusieron comentarios abiertos y evitando la repetición de personas. La descripción de los datos que quedaron se muestra en la tercera columna de la Tabla 2; el porcentaje final de profesores que cumple con estas características se puede observar en la última columna de la misma tabla.

**Tabla 2.** Conjunto de datos para los experimentos.

	Número de profesores sin filtrado	Número de profesores con filtrado	Porcentaje
<b>Bajo</b>	1,654	994	60%
<b>Medio</b>	892	587	66%
<b>Alto</b>	397	277	70%
<b>Total</b>	2,943	1,858	

Partimos del uso de los datos anteriormente descritos para determinar el comportamiento en el empleo de  $n$ -gramas como mecanismo de representación

de los comentarios abiertos realizados por los profesores, y con la finalidad de evaluar la exactitud de un modelo basado en aprendizaje automático para clasificar a los académicos universitarios de acuerdo a su nivel de Burnout. En la siguiente sección se explica cada uno de los experimentos llevados a cabo en este trabajo de investigación.

### 3. Resultados experimentales

En este trabajo se plantean dos experimentos: 1) la evaluación del rendimiento de un clasificador supervisado cuando se utilizan  $n$ -gramas como mecanismo de representación textual. En particular, se varía la longitud del  $n$ -grama, para  $n = 1$  hasta  $n = 5$ , es decir, desde unigramas hasta quintigramas. Este experimento utiliza el conjunto de datos presentado en la sección anterior. La cantidad de profesores por clase, así como la distribución de cada clase se muestra en la Tabla 3. Como puede observarse, la clase “Bajo” tiene una mayor cantidad de datos (53% del conjunto de prueba), superando significativamente a la clase “Alto”, la cual representa únicamente el 15% de los datos del conjunto de prueba.

**Tabla 3.** Conjunto de datos usado en los experimentos.

Nivel de Burnout	Número de profesores	Porcentaje
Bajo	994	53%
Medio	587	32%
Alto	277	15%
Total	1,858	100%

#### 3.1. Resultados obtenidos

Dado que los datos a clasificar son textos de escritura libre, es decir, sin ningún tipo de formato, es necesario realizar una transformación que permita modelar matemáticamente los conjuntos o clases. Hemos elegido un método de representación vectorial (bolsa de palabras) que utilice como característica a un  $n$ -grama. Dado que los datos son discretos, hemos optado por Naïve Bayes Multinomial como el algoritmo de clasificación automático. El componente multinomial puede parecer problemático en el contexto la clasificación de documentos, sin embargo, algunos autores [5] han discutido dichos problemas e incluso han propuesto diversas maneras para aliviarlos, incluyendo el uso de mecanismos de pesado para las características (por ejemplo, tf-idf) en lugar de utilizar solo las frecuencias y la normalización de la longitud del documento, lo cual ha producido un clasificador bayesian que es competitivo con, por ejemplo, máquinas de soporte vectorial.

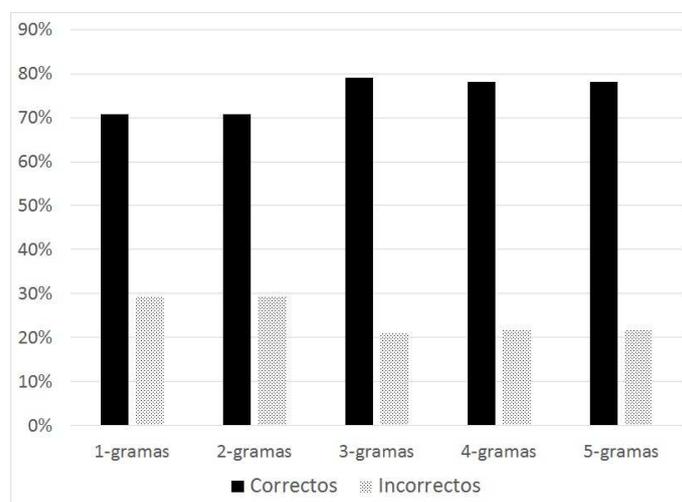
Para el primer experimento realizado se utilizó un esquema de experimentación basado en 10-fold cross-validation, lo cual significa que parte del conjunto de

datos fue utilizado como entrenamiento y el resto como prueba. Se realizan 10 iteraciones y se reporta el promedio de los resultados. Usando el conjunto de datos que contiene 3 clases (ver Tabla 3) se obtienen los resultados mostrados en la Tabla 4. Cada columna reporta el valor de datos clasificados correcta e incorrectamente para  $n$ -gramas ( $n = 1, 2, 3, 4, 5$ ), junto con su valor de exactitud en términos porcentuales.

**Tabla 4.** Comparación de resultados obtenidos con representación basada en  $n$ -gramas ( $n = 1, 2, 3, 4, 5$ ).

	1-gramas	2-gramas	3-gramas	4-gramas	5-gramas
<b>Correctos</b>	1,154/54.05%	1,174/54.99%	1,267/59.34%	1,274/59.67%	994/53.50%
<b>Incorrectos</b>	981/45.95%	961/45.01%	868/40.66%	861/40.33%	864/46.50%

En la Figura 1 se muestra gráficamente el incremento en la exactitud del clasificador cuando la longitud del  $n$ -grama se incrementa desde  $n = 1$  hasta  $n = 4$ , comenzando un decremento cuando el valor de  $n$ -grama alcanza  $n = 5$ . Desafortunadamente, el máximo valor de exactitud es 59.67%, lo cual se considera aún como un resultado pobre para tareas de clasificación en ambientes reales.



**Fig. 1.** Exactitud obtenida al clasificar usando  $n$ -gramas ( $n = 1, 2, 3, 4, 5$ ).

Debido a la problemática anteriormente planteada, se ha decidido verificar el comportamiento del mismo clasificador cuando se utilizan únicamente las dos clases extremas (“Alto” y “Bajo”). El conjunto de datos usado en este segundo experimento se puede observar en la Tabla 5.

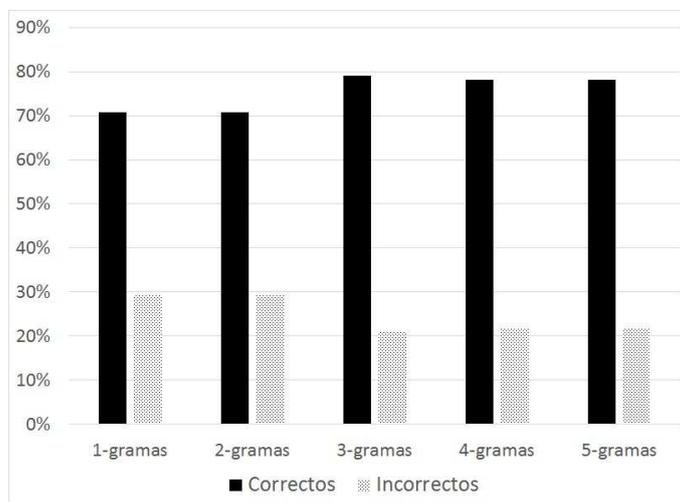
**Tabla 5.** Conjunto de datos usado en los experimentos.

Nivel de Burnout	Número de profesores	Porcentaje
Bajo	994	78.2%
Alto	277	21.8%
<b>Total</b>	<b>1,271</b>	<b>100%</b>

En la Tabla 6 se pueden observar los resultados cuando el algoritmo de clasificación se ejecuta considerando únicamente las dos clases mas alejadas. Se puede ver un incremento considerable en la exactitud de los resultados. El mejor valor obtenido es cuando el tamaño del *n*-grama es 3, con un ligero descenso cuando se incrementa el tamaño del *n*-grama. En la Figura 2 se muestra gráficamente el comportamiento descrito.

**Tabla 6.** Comparación de resultados obtenidos con representación basada en *n*-gramas ( $n = 1, 2, 3, 4, 5$ ).

	1-gramas	2-gramas	3-gramas	4-gramas	5-gramas
<b>Correctos</b>	899/70.73%	890/70.03%	1,005/79.07%	994/78.21%	994/78.21%
<b>Incorrectos</b>	372/29.27%	381/29.97%	266/20.93%	277/21.79%	277/21.79%



**Fig. 2.** Exactitud obtenida al clasificar usando *n*-gramas ( $n = 1, 2, 3, 4, 5$ ).

Finalmente, en la Tabla 7 se enumeran algunos de los patrones de texto encontrados durante el proceso de extracción de características basado en *n*-gramas.

Algunos de estos patrones son evidentemente positivos y estarían asociado a profesores con bajo nivel de Burnout, mientras que otros de éstos son patrones asociados con un alto nivel de Burnout, por ejemplo: “mejorar las condiciones de trabajo”. Un estudio posterior debería determinar la polaridad de dichos patrones y a partir de esta, probablemente generar una relatoría de las posibles mejoras a fin de decrementar el nivel de Burnout de los académicos universitarios.

**Tabla 7.** Patrones encontrados en los comentarios.

<b>Patrón de texto</b>
comunicación directa con la responsable
comunicación entre directivos y docentes
con los compañeros de trabajo
docentes participen en las actividades
el ambiente de trabajo es
no tomar en cuenta a
en mi centro de trabajo
eso impide que haya conflictos
gracias por tomar en cuenta
numero de alumnos por grupo
para la toma de decisiones
por parte de las autoridades
por parte de los directivos
que haya conflictos o diferencias
que se tomen en cuenta
un buen ambiente de trabajo
una mejor relacion
y eso impide que haya
mejorar las condiciones de trabajo

#### 4. Conclusiones

En este trabajo de investigación se ha explorado el proceso de la clasificación automática de comentarios abiertos hechos por profesores, con la finalidad de determinar el nivel de Burnout de dichos académicos universitarios. Se llevaron a cabo dos experimentos, variando la cantidad de clases o niveles de Burnout. Para cada experimento se utilizó una representación basada en  $n$ -gramas, la cual se evaluó desde longitudes  $n = 1$  hasta  $n = 5$ .

Los mejores resultados que se reportan son de un 79.07% de exactitud, cuando se usan únicamente dos clases (nivel “Bajo” y “Alto”), lo cual significa que el modelo generado permite clasificar correctamente 8 de 10 comentarios abiertos y asociar esa clasificación para determinar si el profesor que expresó el comentario se encuentra con uno de estos dos posibles niveles de Burnout.

El análisis final permitió encontrar una serie de patrones textuales que refieren a aspectos académicos pero más laborales y que estarían asociados con

cierto nivel de Burnout. Es trabajo a futuro el desarrollar métodos computacionales que exploten estos patrones para apoyar mejor la tarea de disminuir el nivel de estrés en los académicos universitarios.

## Referencias

1. Freudenberger, H.J.: Staff burn-out. *Journal of Social Issues* 30(1), 159–165 (1974), <http://dx.doi.org/10.1111/j.1540-4560.1974.tb00706.x>
2. Maslach, C., Jackson, S.E.: The measurement of experienced burnout. *Journal of Organizational Behavior* 2(2), 99–113 (1981), <http://dx.doi.org/10.1002/job.4030020205>
3. Maslach, C., Schaufeli, W.B., Leiter, M.P.: Job burnout. *Annual Review of Psychology* 52(1), 397–422 (2001)
4. Olivares, V.E. y Gil-Monte, P.: Análisis de las principales fortalezas y debilidades del maslach burnout inventory (mbi). *Ciencia & Trabajo* 33, 160–167 (2009)
5. Rennie, J.D.M., Shih, L., Teevan, J., Karger, D.R.: Tackling the poor assumptions of naive bayes text classifiers. In: *In Proceedings of the Twentieth International Conference on Machine Learning*. pp. 616–623 (2003)
6. Selye, H.: A syndrome produced by diverse nocuous agents. *Nature* 138(32) (1936)