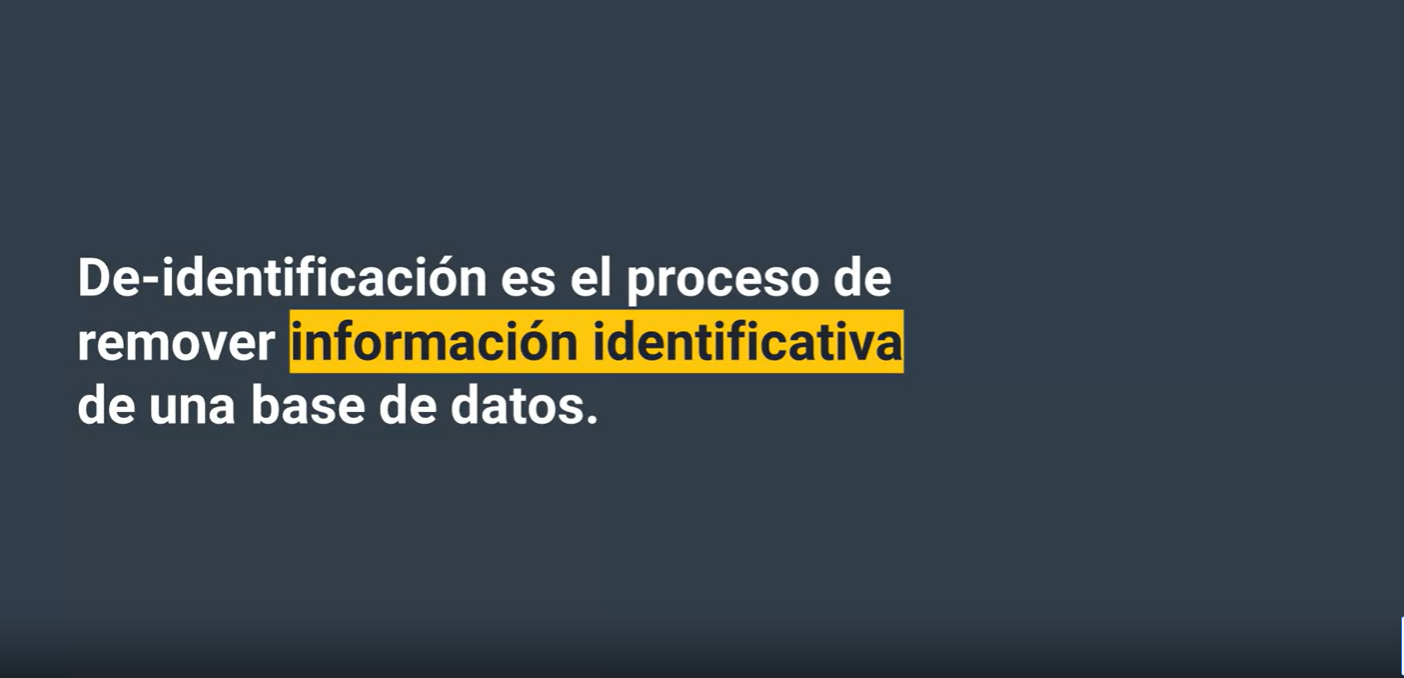
**SEMANA 1**





Tabla

Descripción generada automáticamente

FORO:

Hace un par de años liderada el equipo de BI que se encargaría de generar insights para la nueva estrategia de transformación digital que estaba implementando la compañía en la cual trabajaba. La iniciativa de transformación digital era el principal proyecto de compañía. Se trataba de automatizar a través de 5 canales digitales (Whatsapp, Chatbot, una Aplicación móvil, el IVR y un nuevo portal web), la atención de un millón de llamadas por año que en ese momento se atendían a través de operadores humanos en el call center. Un proyecto que tenia un CAPEX de $2M USD y un plazo de ejecución de 3 años, con medición de avances y de hitos semestral.

Dentro de ese macro-proyecto, la organización formó “tribus”, que se encargarían de abordar distintas iniciativas en pro de ese macro-proyecto. Mi equipo de BI desplegaba capacidades como desarrolladores en el modelo Scrum a distintas tribus, con objetivos diferentes. Cada una de estas tribus se enfrentaba a desafíos muy complejos, especialmente en la parte tecnología y metodológicamente. Era la primera vez que esta compañía trabajaba bajo metodologías ágiles, y fue necesario contratar una empresa de consultoría que le enseñara a todas las personas que estaban en esas tribus cómo trabajar bajo la metodología. Así mismo, se capacitaron unos coach de las metodologías para que se asegurasen el buen uso de este.

Recuerdo particularmente un gran desafío en la tribu de WhatsApp, donde toda la información que se extraía del chatbot estaba en formato JSON. Sin embargo, nadie del equipo era experto en bases de datos No estructuradas. Para solucionarlo, tuvimos la necesidad de contratar un experto en el tema, así como herramientas para manejo de este tipo de información (Bases de datos NO SQL). Por supuesto esto supuso un sobrecosto y un incumplimiento de las fechas pactadas. El hito correspondiente a la extracción y almacenamiento de información logró llegar a buen término después de 5 meses de retraso y un sobrecosto cercado a los 14.000 USD.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1. Se incluye un resumen que destaca la intención y principales puntos del texto:

En el artículo, el autor analiza los problemas recientes con páginas gubernamentales como HealthCare.gov, Sam.gov y USAJobs.gov, los cuales reflejan problemas más importantes con la forma en que el gobierno de EE. UU. compra tecnología. El autor argumenta que la Regulación Federal de Adquisiciones, que rige los contratos gubernamentales, favorece a las empresas que pueden navegar las regulaciones por encima de aquellas que pueden hacer el mejor trabajo.

Por primera vez, un presidente de EE. UU. tuvo que disculparse por un sitio web roto. Standish Group, una firma de investigación asegura que el 94% de los grandes proyectos federales de tecnología de la información en los últimos diez años no tuvieron éxito. Más de la mitad de todos los proyectos se retrasaron, excedieron el presupuesto o no cumplieron con las expectativas de los usuarios, y el 41,4 % fracasó por completo.

Así mismo, se destacan modelos exitosos como el del gobierno británico, quien creó un Servicio digital gubernamental para construir o encontrar la tecnología adecuada para las necesidades del gobierno. Es un equipo de tecnólogos internos cuyo trabajo es construir la tecnología adecuada o encontrar los proveedores adecuados para cada necesidad en todo el gobierno.

Otro aspecto importante en el artículo es poner fin a las grandes compras federales de tecnología de la información y trabajar con pequeñas empresas para facilitar la contratación con el gobierno. Argumentan que cualquier metodología con un 94 % de probabilidad de falla o retraso, que les cuesta a los contribuyentes miles de millones de dólares, no pertenece a un gobierno del siglo XXI.

En conclusión, el autor destaca la necesidad de un gobierno más participativo e interactivo y sugieren que el presidente debe usar el poder de la Casa Blanca para brindar a los tecnólogos expertos de su administración la capacidad de trabajar con las agencias para tomar las decisiones correctas. El artículo insta a los Estados Unidos a encontrar una solución al proceso de adquisición federal que evite que los proyectos de tecnología del gobierno se lastimen a sí mismos al firmar grandes contratos cuyos términos no se cumplen de manera repetida y espectacular.

1. Se realiza un análisis de los principales puntos del texto

Como principales puntos del texto podemos analizar un problema que, aunque pareciera ser exclusivo de países en vía de desarrollo, impacta las grandes potencias. El exceso de burocracia en el sector público. Si pensamos en los EE.UU, tenemos que pensar en Apple, Microsoft, Google, Silicon Valley, compañías que han revolucionado el mundo gracias a los avances tecnológicos. Resulta inverosímil pensar que un país que es fuente de probablemente los mayores desarrollos tecnológicos de los últimos 50 años presente una tasa de fracaso del 94% en proyectos federales de TI. Esta burocracia mal aplicada, termina dilapidando millones de dólares de los contribuyentes, que sueñan y aspiran a tener plataformas que simplifiquen y mejoren su calidad de vida.

Es importante tener en cuenta que la burocracia no es necesariamente algo malo en sí mismo, sino que puede ser una herramienta importante para garantizar la transparencia y la rendición de cuentas en la administración pública. Lo importante es asegurarse de que la burocracia sea eficiente, efectiva y se utilice para el bienestar de la sociedad en su conjunto.

El problema se genera cuando existe un exceso de burocracia que privilegia quien sortea la regulación y no necesariamente garantiza el mejor exponente técnico. Ejemplos como el de Reino Unido sientan un procedente de cómo los países deben abordar este tipo de inconvenientes para garantizar el éxito de los proyectos de TI.

1. El análisis incluye una reflexión personal del estudiante

En Latinoamérica, y en muchos países en vía de desarrollo luchamos con peores niveles de burocracia comparado con economías desarrolladas como la estadounidense, la alemana y la británica, con el agravante que en muchos países, adicionalmente dicha burocracia viene acompañada de una alta tasa de corrupción. En Colombia hace un par de años tuvimos un caso muy renombrado, donde se demostraron manejos irregulares en la contratación pública de tecnologías de conectividad e implementación de cobertura de internet en escuelas de zonas rurales. Un escándalo que salpicó altos funcionarios del gobierno nacional y en el cuál se perdieron alrededor de USD 150.000.

Esto simplemente refuerza la idea que, en estas latitudes, se requieren esfuerzos aun mayores porque luchamos con 2 enemigos muy poderosos para lograr avances significativos tecnológicamente: la burocracia y la corrupción.

1. Se incluye una conclusión

Como profesionales que trabajamos o estudiamos para desarrollar habilidades en ciencia de datos, es importante siempre tener en mente cómo vamos a aportar a un mejor país. Enteniendo las realidades mencionadas anteriormente, tenemos que identificar la forma de aportar al desarrollo tecnológico, sin importar si estamos en el sector público o privado. Para ello, sugiero los siguientes elementos:

1. Identificar los principales desafíos: En este análisis nos centramos en 2 desafios, la burocracia y corrupción. Sin embargo, existen otros elementos a resolver si queremos ver desarrollo tecnológico que acompañen el crecimiento del país desde el sector público. Factores como la falta de acceso a internet en áreas rurales, la falta de políticas de ciberseguridad adecuadas, el alto costo en la adquisición de tecnologías.
2. Establecer una estrategia clara: Una vez identificados los principales desafíos, se debería establecer una estrategia clara para abordarlos. Esta estrategia debería tener objetivos claros y medibles, y ser flexible para adaptarse a los cambios en el entorno político y económico del país.
3. Fomentar la transparencia y la responsabilidad: Para abordar la corrupción, se debería promover la transparencia y la responsabilidad en la adquisición y contratación de tecnologías. Esto se podría lograr a través de la implementación de sistemas de monitoreo y evaluación rigurosos, así como de la participación de la sociedad civil en el proceso de toma de decisiones.
4. Capacitación técnica: La falta de capacitación técnica es otro de los principales desafíos que se enfrentan en la implementación de tecnologías de la información en el sector público. Para abordar este desafío, se podrían establecer programas de capacitación técnica para los empleados públicos, así como para la sociedad en general.
5. Alianzas con el sector privado: Finalmente, se debería fomentar la colaboración con el sector privado para aprovechar su experiencia y recursos en el desarrollo de tecnologías de la información. Esto incluiría la colaboración en la adquisición y contratación de tecnologías, así como en la capacitación técnica y el desarrollo de políticas de ciberseguridad.

En resumen, las acciones propuestas se centrarían en identificar los principales desafíos, establecer una estrategia clara y flexible, fomentar la transparencia y la responsabilidad, brindar capacitación técnica y establecer alianzas con el sector privado para lograr un mayor desarrollo de tecnologías de la información en el sector público.

Semana 3:

Pregunta 1 Sea X una variable aleatoria que representa el lanzamiento de un dado (de 6 caras) ‘justo’; es decir, la probabilidad de que salga cualquiera de sus caras es 1/6. ¿Cuánto es E[X]?

El valor esperado de una variable aleatoria es la suma ponderada de los posibles resultados, donde los pesos son las probabilidades de que ocurran los respectivos resultados. En el caso de lanzar un dado justo, cada uno de los seis resultados posibles tiene una probabilidad igual de 1/6. Por lo tanto, el valor esperado de la variable aleatoria X es:

E[X] = (1/6)(1) + (1/6)(2) + (1/6)(3) + (1/6)(4) + (1/6)(5) + (1/6)(6) = 3.5

Por lo tanto, el valor esperado de X es 3.5, lo que significa que en promedio se espera obtener un resultado de 3.5 al lanzar un dado justo.

Sea X~Ber(1/3). ¿Cuánto es la varianza de X?

La varianza de una variable aleatoria X se define como la esperanza del cuadrado de la desviación de X respecto a su esperanza. Es decir:

Var(X) = E[(X - E[X])^2]

En el caso de una distribución Bernoulli, la esperanza de X es p y la varianza es p(1-p). Por lo tanto, en este caso:

E[X] = 1/3 Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2

Para calcular E[X^2], necesitamos la distribución de probabilidad de X^2:

| **X^2** | **P(X^2)** |
| --- | --- |
| 0 | 2/3 |
| 1 | 1/3 |

Entonces:

E[X^2] = (0)(2/3) + (1)(1/3) = 1/3

Por lo tanto:

Var(X) = E[X^2] - (E[X])^2 = 1/3 - (1/3)^2 = 2/9

Por lo tanto, la varianza de X es 2/9.

Sea X~Ber(1/3) e Y una variable aleatoria discreta tal que: P(Y=0)=2/5 y P(Y=3)=3/5. ¿Cuánto es E[X+Y]?

Y=3\*Z, con Z~Ber(3/5). Por ende, E[X+Y]=E[X]+E[Y]=1/3 + E[3\*Z]=1/3 + 3\*E[Z] = 1/3 + 3\*3/5 = 32/15

Tienes razón, la esperanza de Y es E[Y] = E[3Z] = 3E[Z] = 3(3/5) = 9/5. Por lo tanto, la esperanza de X+Y es:

E[X+Y] = E[X] + E[Y] = 1/3 + 9/5 = 32/15

¡Gracias por la corrección!

me explicas por qué E[3\*Z]=1/3 ?

Tienes razón, hay un error en mi respuesta anterior. La esperanza de 3Z no es igual a 1/3, sino que es igual a 3E[Z], ya que la constante 3 se puede sacar fuera de la esperanza.

En este caso, Z sigue una distribución Bernoulli con parámetro p=3/5, por lo que E[Z] = p = 3/5. Entonces:

E[3Z] = 3E[Z] = 3(3/5) = 9/5

Lamento la confusión y gracias por señalar el error.

espera, otra vez

Claro, ¿en qué puedo ayudarte? ¿Hay alguna otra pregunta o aclaración que necesites? Estoy aquí para ayudarte en lo que necesites.

no entiendo por qué en esta pregunta: Sea X~Ber(1/3) e Y una variable aleatoria discreta tal que: P(Y=0)=2/5 y P(Y=3)=3/5. ¿Cuánto es E[X+Y]? la respuesta dice que Y=3\*Z, con Z~Ber(3/5). Por ende, E[X+Y]=E[X]+E[Y]=1/3 + E[3\*Z]=1/3 + 3\*E[Z] = 1/3 + 3\*3/5 = 32/15. Me explicas?

Claro, lo que se está haciendo en la respuesta es utilizar la regla de la esperanza condicional para encontrar la esperanza de Y. En particular, dado que se sabe que Y toma los valores 0 y 3 con probabilidades 2/5 y 3/5 respectivamente, se puede escribir:

E[Y] = E[Y | Y=0]P(Y=0) + E[Y | Y=3]P(Y=3)

Ahora bien, dado que se asume que Y es una variable aleatoria que toma únicamente los valores 0 y 3, se puede escribir Y = 3Z, donde Z es una variable Bernoulli con parámetro p = 3/5 que toma los valores 0 y 1 con probabilidades 2/5 y 3/5 respectivamente. Así, se tiene que:

E[Y | Y=0] = E[3Z | Y=0] = 3E[Z | Y=0] = 3(0) = 0 E[Y | Y=3] = E[3Z | Y=3] = 3E[Z | Y=3] = 3(1) = 3

Donde la segunda igualdad se sigue de que Y=0 implica que Z=0 y Y=3 implica que Z=1, y la tercera igualdad se sigue de que la esperanza condicional de una variable Bernoulli es igual al parámetro de la distribución condicionada. Entonces, se tiene:

E[Y] = E[Y | Y=0]P(Y=0) + E[Y | Y=3]P(Y=3) = 0(2/5) + 3(3/5) = 9/5

Por lo tanto, la esperanza de X+Y es:

E[X+Y] = E[X] + E[Y] = 1/3 + 9/5 = 32/15

Espero que esto te ayude a entender mejor la respuesta. Si tienes alguna otra pregunta o duda, no dudes en preguntar.

Considere una base de datos S, la cual solo posee las respuestas binarias (1 para SÍ y 0 para NO) de dos personas. Supongamos, por ejemplo, que S=(1,0) ; es decir, la primera persona respondió que SÍ y la segunda que NO. Considere el mecanismo M (que es aleatorizado) que al tomar la base de datos S, suma todas las respuestas: por ejemplo, para el conjunto de datos S=(1,0), M(S) = 1 + 0 = 1 En general, si S = (z1, z2), entonces M(S) = z1 + z2 ¿Cuál es el parámetro de privacidad de M? 1 punto M no es diferencialmente privado M es 1-DP M es ln(3)-DP

El parámetro de privacidad de M está dado por el valor mínimo de la cantidad:

ln[Pr[M(S)=m] / Pr[M(S')=m]]

donde la probabilidad se toma sobre todas las posibles bases de datos S y S' que difieren en un solo elemento, y m es cualquier posible resultado de la función de M.

En este caso, dado que M es una función determinística, es decir, que M(S) siempre devuelve la misma salida para cualquier S, entonces la probabilidad de que M(S) tome un valor m es 1 si m = M(S), y 0 en otro caso. Por lo tanto, la expresión anterior se reduce a:

ln[1 / Pr[M(S')=M(S)]]

Para que M sea diferencialmente privado, esta cantidad debe ser menor o igual a ε para cualquier S y S' que difieran en un solo elemento y para cualquier valor posible de M(S). Esto implica que se debe encontrar el valor máximo de Pr[M(S')=M(S)] para cualquier S y S' que difieran en un solo elemento. En este caso, dado que M(S) siempre toma un valor entre 0 y 2, el valor máximo de esta probabilidad es 1/2, que se alcanza cuando S y S' difieren solo en un elemento.

Por lo tanto, se tiene que:

ln[1 / Pr[M(S')=M(S)]] <= ln[1 / (1/2)] = ln(2)

Lo cual implica que M es ln(2)-DP. En particular, M no es 1-DP, ya que su parámetro de privacidad es mayor que 1. Además, dado que M no cumple con la propiedad de la diferencial de Laplace, en general no se puede garantizar que M sea diferencialmente privado para cualquier valor de ε.

Interfaz de usuario gráfica, Aplicación, Teams

Descripción generada automáticamente

Texto, Carta

Descripción generada automáticamente con confianza media

Por definición, un mecanismo es diferencialmente privado si para cualquier par de conjuntos de respuestas posibles $S$ y $S'$ y cualquier resultado posible $r$ del mecanismo, se cumple que:

��[�(�)=�]≤��⋅��[�(�′)=�]*Pr*[*M*(*S*)=*r*]≤*eϵ*⋅*Pr*[*M*(*S*′)=*r*]

donde $M$ es el mecanismo, y $\epsilon$ es el parámetro de privacidad.

En este caso, consideremos dos conjuntos de respuestas posibles $S = {(0,0), (0,1), (1,0), (1,1)}$ y $S' = {(0,1), (1,0), (1,1), (0,0)}$. Sea $r = (1,1)$ el resultado del mecanismo.

Si el primer encuestado responde (0,1), entonces el mecanismo revertirá ambas respuestas y dará como resultado (1,0), por lo que $Pr[M(S) = (1,1)] = 0$. Sin embargo, si el segundo encuestado responde (1,1), el mecanismo no tendrá que revertir ninguna respuesta y dará como resultado (1,1), por lo que $Pr[M(S') = (1,1)] = 1/4$. Por lo tanto:

��[�(�)=(1,1)]=0≰��⋅14=�ln⁡(4)4=44=1*Pr*[*M*(*S*)=(1,1)]=0≰*eϵ*⋅41​=4*e*ln(4)​=44​=1

Para cualquier valor de $\epsilon$, por lo que el mecanismo no satisface la definición de privacidad diferencial.