# Sigurnost u računarskim mrežama

## Filtriranje neželjene elektronske pošte

*Nemanja Maček*

# Sadržaj

- Struktura elektronske pošte
- Problem neželjene elektronske pošte
- Jednostavne metode filtriranja
- Klasifikacija tekstualnih dokumenata
- Bajesovo filtriranje
- Metoda najbližih suseda
- Obrada korpusa
- Izabrani napadi na filtre neželjene elektronske pošte
- Pecanje

# Zanimljivost

**@**.

- U oktobru 1971. (mada ima izvora koji tvrde da se to dogodilo jula 1970.) Rej Tomlinson je napisao prvi program za razmenu poruka između dva računara.
- Pored programa za razmenu e-pošte, Rej Tomlinson je odgovoran za masovnu upotrebu znaka @ (“et”, poznato i kao “ludo a” ili “majmunsko a”).
- Smišljajući kako da razvrsta primaoce poruka odlučio je da njihova imena i imena računara na kojima se nalaze njihovi nalozi razdvoji nekim znakom interpunkcije.
- Kako je na svojoj tastaturi imao samo 12 takvih znakova na raspolaganju, odlučio se za onaj koji se nikada ne koristi u pisanju poruka*.
- Taj princip adresiranja koristi se i danas.

* Tom Van Vleck: The history of electronic mail. http://www.multicians.org/thvv/mail-history.html. [Online; pristupljeno 16 Nov. 2015].

# Struktura elektronske pošte

**Telo i zaglavlje poruke e-pošte**.

- **Elektronska pošta** je mrežni servis koji omogućava slanje i primanje poruka raznovrsnog sadržaja.
- Svaka poruka elektronske pošte se sastoji iz zaglavlja i tela poruke.
- **Zaglavlje** se sastoji od strukturiranih podataka kao što su:
  - adresa i ime pošiljaoca,
  - adresa primaoca,
  - adresa servera preko kojeg je poruka poslata, kao i servera koji su prosleđivali poruku na njenom putu do odredišta,
  - datum slanja,
  - ime programa koji je korišćen za slanje poruke i
  - prioritet.

# Struktura elektronske pošte

**Telo i zaglavlje poruke e-pošte**.

- Svaka stavka zaglavlja se sastoji od imena stavke i vrednosti stavke.
- Sledi primer jednog dela zaglavlja poruke:

Date: Thu, Nov 16, 2017 at 3:22 PM
To: nemanja.macek@acme.com
From: pera.kojot.supergenije <p.kojot@acme.com>
Reply-to: ptica.trkacica@acme.com
mailed-by: acme.com
Subject: Naslov poruke
X-Priority: 3

# Struktura elektronske pošte

**Telo i zaglavlje poruke e-pošte**.

- **Telo** se sastoji od nestrukturiranih podataka kao što je tekst, HTML, multimedijalni sadržaj i prilozi.
- Dakle, telo poruke se može sastojati iz više delova u zavisnosti od toga da li se sa tekstom poruke šalju i datoteke.
- Ukoliko se šalju onda se u samom zaglavlju poruke to može označiti pomoću stavke sledećeg oblika:

```
Content-Type: multipart/alternative;
boundary="b1_8d5c3f4ac4f174c9d0bbc13814d16891"
```

# Problem neželjene elektronske pošte

**Šta je pretnja korisnicima e-pošte**?

- Korišćenje e-pošte je ugroženo sledećim “pojavama”:
  - bombardovanje porukama,
  - neželjena elektronska pošta,
  - pokušaji preuzimanja ličnih podataka i
  - prenošenje zlonamernog softvera.

# Problem neželjene elektronske pošte

- **Šta je *spam*?**
  - *Spam* predstavlja zloupotrebu elektronskih sistema u svrhu slanja neželjenih masovnih poruka bez ikakvog kriterijuma.
  - **Vrste *spam*-a**:
    - čet spam,
    - Veb spam
    - SMS spam,
    - forum spam,
    - spam na društvenim mrežama,
    - e-mail spam (neželjena elektronska pošta), itd.

# Problem neželjene elektronske pošte

- **Šta je *spam* u kontekstu e-pošte**?
  - Na Internetu termin *spam* označava nepoželjnu, besciljnu (engl. *untargeted*) elektronsku poštu.
  - U osnovi, to je slanje komercijalnih poruka, najčešće reklama i mrežnih marketing šema, na stotine hiljada pa i miliona adresa korisnika širom mreže, bez njihovog odobrenja.
  - Blaži oblik spama predstavlja TDEM (engl. *Targeted Direct Email Marketing*) ili direktni marketing putem ciljnih grupa.

# Problem neželjene elektronske pošte

- **Šta ne spada u neželjenu e-poštu**?
  - Tipičan primer – sami ste se pretplatili na neki servis.
  - Ukoliko postoji mogućnost da otkažete primanje ovakvih prouka, onda se ove poruke ne smatraju za neželjenu e-poštu.

You are receiving this e-mail as a part of your subscription to NorwegianClass101.com

This email is intended for: [redacted]@gmail.com

**Remove me** from the NorwegianClass101.com mailing list.

Innovative Language Learning
1F Senshu Building 3-4-4 Akasaka - Minato-ku, Tokyo 107-0052

# Problem neželjene elektronske pošte

- **Okvira raspodela neželjene e-pošte**.
  - Raspodela se menja sa vremenom, ali daje dobar uvid šta se već duže vreme nalazi na prva četiri mesta!
    - Proizvodi – 25%,
    - finansijski – 20%,
    - za odrasle – 19%,
    - prevara – 9%, itd.

# Problem neželjene elektronske pošte

- **Primer: Nigerijska prevara**.
  - Internet je prepun ljudi koji žele da prevare nekog, a jedan od najpopularnijih načina je takozvana Nigerijska prevara*.
  - U pitanju je e-pošta u kojoj piše da ste navodno dobili ogroman novac, iako ni sami ne znate osobu koja vam je to saopštila.
  - Prevara počinje tako što vas nepoznate osobe kontaktiraju putem e-pošte nudeći ogromnu sumu novca koji ste navodno dobili ili nasledili.
  - Kako biste dobili nasledstvo ili novac koji žele da vam pošalju, od vas će tražiti da platite razne takse.
  - U početku je to mala količina novca, kako bi sve delovalo uverljivo, ali ukoliko nasednete, sigurno će izmisliti još nešto što mora da se plati kako bi dobili svoju nagradu.

* Prvi talas ovih prevara potekao je iz Nigerije, pa je po tome i dobila ime.

# Zanimljivost

- **Danijel Balsam** iz Amerike je gospodin koji je toliko bio iznerviran zbog stalnih spam mejlova da je dao otkaz, završio parvo (*University of California*, *Hastings College of the Law*) i do današnjeg dana je zaradio preko milion dolara tužeći ljude i kompanije koji su odgovorni za te reklame*.

* P. Elias (December 26, 2010). "Man quits job, makes living suing e-mail spammers (archived)". Associated Press.

# Problem neželjene elektronske pošte

- **Neki saveti za zaštitu od neželjenih poruka**.
  - **Ne otvarajte neželjene poruke**, izbrišite ih bez prethodnog otvaranja.
    - Spam se na neki način povezuje sa prevarama poput pecanja.
    - Postoji takođe i mogućnost da se zarazite nekim zlonamernim softverom.
  - Nikada **ne odgovarajte** na spam poruke bez obzira šta se od vas traži.
    - U suprotnom na ovaj način obaveštavate pošiljaoca o validnosti vaše e-mail adrese.
  - **Ne kupujte proizvode** koji se reklamiraju putem spam poruka bez obzira koliko povoljni bili.
    - Ukoliko kupite proizvod očekujte još više reklamnih poruka.
  - **Ne prosleđujte spam poruke**.
    - Ukoliko nastavite sa prosleđivanjem vaša adresa će stići i kod ostalih spamera.
      - Dobijaćete još više neželjenih poruka.
    - Postoji opasnost da je poruka sadržala zlonamerni kod (na primer, virus), samim tim postajete pošiljalac virusa.

# Jednostavne metode filtriranja

- **Metoda bele liste** (engl. *whitelisting*).
  - Prihvatanje svih poruka e-pošte pristiglih sa adresa koje se nalaze na beloj listi.
  - Sastavni deo većine implementacija metoda sive liste.
    - U slučaju da se adresa primljene poruke nalazi na beloj listi, poruka se odmah dostavlja, čime se izbjegavaju kašnjenja uzrokovana analizom pomoću metode sive liste.
- **Metoda crne liste** (engl. *blacklisting*).
  - Koristi se popisom adresa s kojih je u određenom proteklom periodu pristigla e-pošta koja je klasifikovana kao *spam*.
  - Iako se popis može čuvati lokalno, najčešće se ti popisi proveravaju u realnom vremenu, sa servera namenjenih upravo tome.
  - Takvi serveri se često ažuriraju i u svakom trenutku sadrže crne liste trenutno aktivnih pošiljalaca spam poruka.

# Jednostavne metode filtriranja

- **Metoda sive liste** (engl. *graylisting*).
  - Prilikom pokušaja dostavljanja poruke e-pošte metoda sive liste pregleda **triplet**:
    - **IP adresu računara** s kog je poslata poruka, **adresu pošiljaoca** i **adresu primaoca**.
  - U slučaju da je određeni triplet prvi put viđen odbija se njegova isporuka.
  - Protokol SMTP specificira mogućnost **privremene nemogućnosti isporuke** elektronske pošte.
  - Valjani server elektronske pošte – MTA (*Mail Transfer Agent*) nakon određenog vremenskog intervala **pokušava da ponovi isporuku**.
  - Ova je činjenica bitna!
    - Većina spam poruka šalje pomoću aplikacija koje su razvijene samo u tu svrhu.
    - One ne implementiraju u potpunosti SMTP protokol.
      - Ne pokušavaju da ponove isporuku.
    - Najčešće koriste privremene dinamičke IP adrese što automatski onemogućava ponovni pokušaj slanja poruke.

# Klasifikacija tekstualnih dokumenata

- **Definicija**.
  - Intuitivno, klasifikacija teksta je zadatak klasifikovanja dokumenta u jednu ili više predefisanih kategorija.
  - Formalnije, ako je $d_i$ dokument iz skupa dokumenata $D$ i $\{c_1, c_2, \dots, c_n\}$ skup svih kategorija, tada je klasikacija teksta dodeljivanje jedne ili više kategorija $c_j$ dokumentu $d_i$.
- **Formalna definicija kategorizacije teksta**.
  - Kategorizacija teksta je zadatak dodeljivanja bulovskih vrednosti svakom paru $\langle d_i, c_j \rangle \in \mathcal{D} \times \mathcal{C}$, gde je $\mathcal{D}$ domen dokumenata i $\mathcal{C} = \{c_1, c_2, \dots, c_{|\mathcal{C}|}\}$ skup predefinisanih kategorija.
    - Vrednost T (*True*) dodeljena paru $\langle d_i, c_j \rangle$ ukazuje na odluku da je dokument $d_i$ pod kategorijom $c_j$.
    - Vrednost F (*False*) dodeljena paru $\langle d_i, c_j \rangle$ ukazuje na odluku da dokument $d_i$ nije pod kategorijom $c_j$.

# Klasifikacija tekstualnih dokumenata

- **Jednoznačna i višeznačna klasifikacija**.
  - Različite zavisnosti se mogu javiti pri klasifikaciji teksta u zavisnosti od primene.
  - Na primer, za dato celobojno $k$ tačno $k$ elemenata iz $\mathcal{C}$ treba dodeliti svakom $d_i \in \mathcal{D}$.
    - Slučaj u kome tačno jedna kategorija mora biti dodeljena svakom dokumentu $d_i \in \mathcal{D}$ se naziva **jednoznačna** (engl. *single-label*) klasifikacija.
    - Slučaj u kome se svakom dokumentu $d_i \in \mathcal{D}$ dodeljuje više od jedne kategorije naziva se **višeznačna** (engl. *multi-label*) klasifikacija.
      - Specijalan slučaj jednoznačne klasifikacije je binarna klasifikacija u kojoj se svakom dokumentu $d_i \in \mathcal{D}$ dodeljuje kategorija $c_j$ ili njen komplement $\overline{c_j}$.
  - Teorijski, binarni slulaj (jednoznačna klasifikacija takođe) je opštiji od višeznačne, jer se algoritam binarne klasifikacije može upotrebiti i za višeznačnu klasifikaciju.
    - Kategorije $\{c_1, c_2, \dots, c_n\}$ se podele u $|\mathcal{C}|$ nezavisnih problema binarne klasifiacije kao $\{c_j, \overline{c_j}\}$ za $i = 1, \dots, |\mathcal{C}|$.
    - Potrebno je da su kategorije **stohastički nezavisne** jedna od druge.

# Klasifikacija tekstualnih dokumenata

- **Ocena kvaliteta klasifikatora teksta**.
  - Tabela prikazuje matricu konfuzije za problem binarne klasifikacije.
  - Svaki unos $f_{ij}$ u ovoj tabeli označava broj zapisa iz klase $i$ koji su predviđeni u klasi $j$.
    - Na primer, $f_{01}$ je broj zapisa iz klase $0$ pogrešno predviđenih kao klasa $1$.
  - Na osnovu unosa iz matrice konfuzije ukupan broj:
    - ispravnih predviđanja koje je napravio model je $(f_{11} + f_{00})$, a
    - ukupan broj neispravnih predviđanja je $(f_{10} + f_{01})$.

| | | Predviđena klasa | | |
|---|---|---|---|---|
| | | Klasa = 1 | Klasa = 0 | Ukupno |
| Stvarna klasa | Klasa = 1 | $f_{11}$ | $f_{10}$ | $f_{11} + f_{10}$ |
| | Klasa = 0 | $f_{01}$ | $f_{00}$ | $f_{01} + f_{00}$ |
| | Ukupno | $f_{11} + f_{01}$ | $f_{10} + f_{00}$ | $N$ |

# Klasifikacija tekstualnih dokumenata

- **Ocena kvaliteta klasifikatora teksta**.
  - Vrednosti matrice konfuzije obrazložene na primeru klasifikacije elektronske pošte:
    - $f_{11}$ – vrednost “stvarno pozitivni” (engl. *true positive*) predstavlja broj dokumenata koji su zaista spamovi a koje je klasikator prepoznao kao spamove.
    - $f_{01}$ – vrednost “lažno pozitivni” (engl. *false positive*) predstavlja broj dokumenata koji nisu spamovi a koje je klasifikator klasifikovao u grupu spamova.
    - $f_{00}$ – vrednost “stvarno negativni” (engl. *true negative*) predstavlja broj dokumenata koji nisu spamovi a klasifikator ih je klasifikovao kao legitimne poruke.
    - $f_{10}$ – vrednost “lažno negativni” (engl. *false negative*) predstavlja broj dokumenata koji su spamovi a koji su svrstani u grupu legitimnih poruka.

# Klasifikacija tekstualnih dokumenata

- **Ocena kvaliteta klasifikatora teksta**.
  - **Tačnost** (engl. *accuracy*) klasifikatora je mera koja nam daje procenat uspešno klasifikovanih dokumenata.

$$tačnost = \frac{broj\ ispravnih\ predvidjanja}{ukupan\ broj\ predviđanja} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} = \frac{TN + TP}{TP + FN + FP + TN}$$

  - Za mnoge primere klasifikacije tačnost je korisna mera.
  - Međutim, postoje slučajevi kada nam tačnost ne može dati informacije koje tražimo.
    - To su uglavnom scenariji u kojima je jedna klasa značajno manja od druge.
  - Ekvivalentno, performanse modela se mogu izraziti u obliku **stope greške**.

$$stopa\ greške = \frac{broj\ pogrešnih\ predvidjanja}{ukupan\ broj\ predviđanja} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} = \frac{FN + FP}{TP + FN + FP + TN}$$

- **Ocena kvaliteta klasifikatora teksta**.
  - **Preciznost** (engl. *precission*) je mera koja nam daje informaciju o udelu stvarno pozitivnih instanci.
    - Od svih poruka koje su označene kao spam, koji procenat čine poruke koje su stvarno spam?

$$P = \frac{f_{11}}{f_{11} + f_{01}} = \frac{TP}{TP + FP}$$

  - **Odziv** (engl. *recall*) je mera suprotna preciznosti.
    - Od svih poruka koje su stvarno spam, koji procenat poruka je klasifikovan kao spam?

$$R = \frac{f_{11}}{f_{11} + f_{10}} = \frac{TP}{TP + FN}$$

- **Ocena kvaliteta klasifikatora teksta**.
  - **F-mera** je parametar koji u obzir uzima i preciznost i odziv sistema:

$$F = 1/\left(\alpha\frac{1}{P} + (1-\alpha)\frac{1}{R}\right), \alpha \in [0,1]$$

  - Za $\alpha < 1/2$, F-mera nadglašava odziv, a za $\alpha = 0$, F-mera je jednaka odzivu, tj. $F = R$.
  - Za $\alpha > 1/2$, F-mera nadglašava preciznost, a za $\alpha = 1$, F-mera je jednaka preciznošću.
  - Za $\alpha = 1/2$, F-mera predstavlja harmonijsku sredinu preciznosti i odziva, i naziva se balansiranom F-merom:

$$F = \frac{2PR}{P+R}$$

  - Za date vrednosti preciznosti i odziva, blansirana F-mera uvek naginje manjoj vrednosti.

- **Ocena kvaliteta klasifikatora teksta**.
  - Primer.
    - Neka su vrednosti preciznosti odziva za hipotetički klasifikator $P = 0{,}01\%$ i $R = 100\%$.
    - U ovom slučaju, balansirana F-mera iznosi $F = \frac{2PR}{P+R} \approx 0{,}2\%$, tj. naginje preciznosti I predstavlja adekvatniju ocenu sistema.
  - Pitanje.
    - Kako bi ste ocenili ovakav filtar (ukoliko u obzir uzmete preciznost i odziv)?

# Klasifikacija tekstualnih dokumenata

- **Ocena kvaliteta klasifikatora teksta**.
  - Tehnika za evaluaciju klasifikatora koja se često koristi je **unakrsna validacija** (engl. *cross validation*).
    - Skup podataka se podeli na *k* delova približno iste veličine, a zatim se svaki od *k*-1 delova koristi kao skup za učenje a sam taj preostali deo kao skup za testiranje.
    - Postupak se ponavlja *k* puta tako da je svaki od delova po jednom učestvovao u ulozi testnog skupa podataka.
    - Greška klasifikacionog modela je **prosečna greška** svih *k* iteracija u postupku.
  - Često se postupak slučajnog izbora modifikuje tako da se osigura približno jednaka zastupljenost klasa u svakom od k delova.
    - Taj postupak se naziva **stratifikacija**.
  - Složenost evaluacije klasifikacionog modela ovom metodom zavisi od *k*.
  - U praksi se najčešće uzima $k = 5$ ili $k = 10$ jer se takva unakrsna validacija pokazala kao dovoljno tačna a nije previše zahtevna.

# Klasifikacija tekstualnih dokumenata

- **Jednostavan primer obuke**.
  - Na osnovu trening podataka izvode se klasifikaciona pravila za klasifikator.
  - Trening instance:
    - kupovina odmah $\rightarrow$ SPAM,
    - kupovina proizvoda $\rightarrow$ SPAM,
    - dobra kupovina $\rightarrow$ SPAM,
    - dobra plata $\rightarrow$ LEGITIMNA,
    - dobra večera $\rightarrow$ LEGITIMNA,
  - Izvedeno klasikaciono pravilo:
    - kupovina $\rightarrow$ SPAM

# Bajesovo filtriranje

- **Neformalno objašnjenje**.
  - Najpopularnija tehnika filtriranja e-pošte je Bajesovo filtriranje, koje predstavlja jednostavan metod koji se zasniva na verovatnoći.
  - U kontekstu e-pošte to znači da se verovatnoća da je pošta neželjena može izračunati na osnovu određenih reči, ukoliko znamo verovatnoću neželjenih poruka i podatak koliko često se te reči pojavljuju u njima, a koliko u telima sve e-pošte.
  - Različite reči imaju u opštem slučaju različite verovatnoće pojavljivanja u neželjenoj i legitimnoj e-pošti.

# Bajesovo filtriranje

- **Neformalno objašnjenje**.
  - Većina korisnika kada vidi reć Viagra zna da se radi o neželjenoj pošti.
  - Filtar ne zna ove verovatnoće unapred i neophodan mu je trening da bi mogao da odredi verovatnoće pojavljivanja određenih reči u neželjenoj pošti.
    - Treniranje filtra se svodi na to da korisnici ručno obeležavaju da li je pošta nećeljena ili ne.
    - Za sve reči svake poruke pri testiranju filtar će prilagoditi verovatnoće u svojoj bazi podataka.
  - Nakon treniranja poznate su verovatnoće reči koje se nalaze u bazi i moguće je izračunati verovatnoću da li je pošta neželjena ili ne.
    - Ukoliko je izračunata verovatnoća svih reči u poruci veća od nekog praga (npr. 95%) filtar će označiti ovu poruku kao neželjenu.

# Bajesovo filtriranje

- **Matematička osnova**.
  - Bajesov filtar e-pošte koristi Bajesovu teoremu:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

  - gde je:
    - $P(A)$ – početna verovatnoća događaja *A*,
    - $P(B)$ – početna verovatnoća pojavljivanja instance *B*,
    - $P(B|A)$ – uslovna verovatnoća pojavljivanja instance *B* uz uslov ispravnosti događaja A,
    - $P(A|B)$ – uslovna verovatnoća ispravnosti događaja *A* nakon pojavljivanja instance *B*.
      - Zanimljiva je sa stanovišta indukcije znanja jer omogućava procenu ispravnosti događaja nakon posmatranja pojave novih instanci *B*.

# Bajesovo filtriranje

- **Matematička osnova**.
  - Kod konačno mnogo disjunktnih slučajeva $A_i, i = 1, \dots, n$ Bajesova teorema glasi:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

# Bajesovo filtriranje

- **Matematička osnova**.
  - Primer Bajesove teoreme.
    - Lekar zna da meningitis u 50% slučajeva prouzrokuje kočenje vrata.
    - Prethodna (poznata) verovatnoća da bilo koji pacijent ima meningitis je 1/50000.
    - Prethodna verovatnoća da bilo koji pacijent ima ukočen vrat je 1/20.
    - Ako pacijent ima ukočen vrat, koja je verovatnoća da ima i meningitis?

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = \frac{0.5 \times \frac{1}{50000}}{\frac{1}{20}} = 0{,}0002 = 0{,}02\%$$

# Bajesovo filtriranje

- **Kako filtar radi**?
  - Pretpostavimo da poruka sadrži reč kupovina.
  - Filtar će koristiti sledeću formulu koja se zasniva na Bajesovoj teoremi:

$$P_r(S|W) = \frac{P_r(W|S) \times P_r\,(S)}{P_r(W|S) \times P_r(S) + P_r(W|H) \times P_r(H)}$$

  - $P_r(S|W)$ – verovatnoća da je poruka neželjena ako znamo da se reč kupovina nalazi u njoj.
  - $P_r(S)$ – ukupna verovatnoća da je bilo koja data poruka neželjena.
  - $P_r(W|S)$ – verovatnoća da se reč kupovina nalazi u neželjenoj poruci.
  - $P_r(H)$ – ukupna verovatnoća da je bilo koja data poruka legitimna.
  - $P_r(W|H)$ – verovatnoća da se reč kupovina nalazi u legitimnoj poruci.

# Bajesovo filtriranje

- **Nepristrasni filtri**.
  - Statistike pokazuju da je verovatnoća da je neka poruka neželjena 80%, pa je: $P_r(S) = 0{,}8$ i $P_r(H) = 0{,}8$.
  - Međutim većina filtara neželjene pošte pravi pretpostavku da ne postoji razlog da dolazeća pošta ima veću verovatnoću da bude neželjena, tj: $P_r(S) = 0{,}5$ i $P_r(H) = 0{,}5$.
  - Za filtre koji koriste ovu hipotezu kaže sa da su **nepristrasni** jer nemaju predrasude o dolazećoj pošti.

# Bajesovo filtriranje

- **Nepristrasni filtri**.
  - Ova pretpostavka dozvoljava nam pojednostavljivanje polazne formule na sledeći način:

$$P_r(S|W) = \frac{P_r(W|S)}{P_r(W|S) + P_r(W|H)}$$

  - Broj $P_r(W|S)$ je aproksimiran frekvenciji poruka koje sadrže reč kupovina u porukama koje su identifikovane kao neželjene u fazi učenja.
  - Broj $P_r(W|H)$ je aproksimiran frekvenciji poruka koje sadrže reč kupovina u porukama koje su identifikovane kao legitimne u fazi učenja.
  - Da bi ove aproksimacije imale smisla neophodno je da skup poruka u fazi učenja bude:
    - dovoljno **velik** i **reprezentativan**,
    - usaglašen sa 50% hipoteze o ponovnoj podeli poruka, odnosno da su skupovi podataka za neželjenu i legitimnu poštu iste veličine.

# Bajesovo filtriranje

- **Nepristrasni filtri**.
  - Napomena:
    - Vrednost $P_r(W|S)$ se u literaturi popularno naziva *spamicity* ili *spaminess* i predstavlja procenat pojavljivanja date reči u neželjenoj pošti.
    - Ona se računa za svaku reč koja se pojavljuje u poruci i kreće se u opsegu od 0 do 1.

# Bajesovo filtriranje

- **Kombinovanje individualnih verovatnoća**.
  - Odlučivanje da li je poruka neželjena ili legitimna na osnovu samo jedne reči (kupovina) je podložno greškama, pa zbog toga Bajesov filtar pokušava da napravi zaključak na osnovu više reči kombinovanjem više verovatnoća.
  - Većina Bajesovih algoritama za filtriranje e-pošte su zasnovani na formulama koje su striktno validne samo ako su reči prezentovane u poruci **uslovno nezavisni događaji**.
  - Ovaj uslov **nije generalno zadovoljen**.
    - Primer: u prirodnim jezicima verovatnoća nalaženja prideva zavisi od verovatnoće javljanja imenica u tekstu.
    - Korisno je kao idealizacija jer su statističke korelacije pojedinih reči obično nepoznate.

# Bajesovo filtriranje

- **Kombinovanje individualnih verovatnoća**.
  - Izvodimo na osnovu prethodno pomenutog i Bajesove teoreme:

$$p = \frac{p_1 p_2 \dots p_N}{p_1 p_2 \dots p_N + (1 - p_1)(1 - p_2) \dots (1 - p_N)}$$

  - $p$ – verovatnoća da je poruka neželjena,
  - $p_1$ – verovatnoća da je poruka neželjena znajući da sadrži reč kupovina, tj. $p(W_1|S)$,
  - $p_2$ – verovatnoća da je poruka neželjena znajući da sadrži reč viagra, tj. $p(W_2|S)$,
  - …
  - $p_N$ – verovatnoća da je poruka neželjena znajući da sadrži *N*-tu reč (jeftino), tj. $p(W_N|S)$.

  - Spam fltri bazirani na ovoj formuli ponekad se nazivaju **naivni Bajesovi klasifikatori**.
  - Rezultat $p$ se tipično poredi sa **pragom odluke**.
    - Ako je $p$ niže od praga poruka se smatra legitimnom, inače se smatra neželjenom.

# Bajesovo filtriranje

- **Rad sa retkim rečima**.
  - U slučajevima kada reč nikada nije korišćena u fazi učenja i brojilac i imenilac su jednaki nuli.
  - Filtar može da odluči da odbaci takve reči za koje ne postoji informacija.
  - Reči sa kojima smo se susreli samo nekoliko puta u fazi učenja prouzrokovaće problem jer je pogrešno slepo verovati informacijama koje pružaju.
  - Jednostavno rešenje je da se izbegava uzimanje u obzir takvih nepouzdanih reći.

# Bajesovo filtriranje

- **Rad sa retkim rečima**.
  - Neki programi koriste sledeću korigovanu verovatnoću:

$$P_r'(S|W) = \frac{s \times P_r(S) + n \times P_r(S|W)}{s + n}$$

  - $P_r'(S|W)$ – korigovana verovatnoća da je poruka neželjena ako sadrži datu reč.
  - $s$ – vrednost koju prosleđujemo kao dodatnu informaciju o dolaznoj neželjenoj poruci.
  - $P_r(S)$ – verovatnoća da je dolazna poruka neželjena.
  - $n$ – broj pojavljivanja date reči tokom faze učenja.
  - $P_r(S|W)$ – verovatnoća da je poruka neželjena ako znamo da se data reč nalazi u njoj.

# Bajesovo filtriranje

- **Druge heuristike**.
  - Neutralne reči poput “*the*”, “*a*”, “*some*” ili “*is*” (na engleskom) ili njihovi ekvivalenti na drugim jezicima mogu da se ignorišu.
  - Neki Bajesovi filtri ignorišu reći čiji je *spamicity* blizu 0,5 jer veoma slabo dobrinose dobrom odlučivanju.
    - Reči koje se uzimaju u razmatranje su one čiji je spamicity blizu 0 (karakteristični znaci da je poruka legitimna) ili 1 (da je poruka neželjena).
  - Neki softverski proizvodi koriste šablone (**nizove reči**) umesto izolovanih reči prirodnih jezika.
    - Na primer, ukoliko koristimo šablone za kontekst od četiri reči, *spamicity* će se računati nad nizom od te četiri reči, poput “vijagra je dobra za”, umesto izračunavanja *spamicity* za odvojene reči “vijagra”, “je”, “dobra” i “za”.
    - Samim tim ovaj metod daje više osetljivosti na kontekst.

# Bajesovo filtriranje

- **Primena**.
  - Kada prihvatimo poruku $m$ moramo definisati funkciju odlučivanja $f$ koja dodeljuje poruku $m$ svojoj klasi.
  - Spam poruke obeležavaćemo sa $S$, a legitimne sa $L$.
  - Ako je $G_M$ skup poruka tada funkciju $f$ definišemo na sledeći način:

$$f: G_M \rightarrow \{S, L\}$$

# Bajesovo filtriranje

- **Primena**.
  - Kod ovakvih tehnika prvo moramo proveriti neke karakteristike koje mogu uticati na klasifikaciju poruke.
  - Pozivaćemo se na te karakteristike korišćenjem vektora $\vec{x}$.
  - Neka je $P(\vec{x}/c)$ verovatnoća da klasa $c$ generiše poruku čiji je **vektor karakteristika** $\vec{x}$.
  - Ako pretpostavimo da:
    - legitimna poruka nikad nije sadržala tekst $t$ = "kupi sada", i da je
    - $x = (m = utv)$, gde su $u$ i $v$ dva stringa,
  - tada je verovatnoća $P(\vec{x}/L) = 0$.
  - Sada je problem izračunati verovatnoću da poruka koja sadrži karakteristični vektor $\vec{x}$ pripada klasi $c$, odnosno: $P(c/\vec{x})$.

# Bajesovo filtriranje

- **Primena**.
  - To možemo dobiti posmatranjem Bajesovog pravila:

$$P(c/\vec{x}) = \frac{P(\vec{x}/c) \times P(c)}{P(\vec{x})} = \frac{P(\vec{x}/c) \times P(c)}{P(\vec{x}/S) \times P(S) + P(\vec{x}/L) \times P(L)}$$

  - $P(\vec{x})$ je apriorna verovatnoća pojavljivanja poruke čiji je karakteristični vektor $\vec{x}$.
  - $P(c)$ je verovatnoća da bilo koja poruka pripada klasi $c$.
  - Znajući verovatnoće $P(c)$ i $P(\vec{x}/c)$ može se zaključiti $P(c/\vec{x})$.
  - **Pravilo klasifikacije**: ako je $P(S/\vec{x}) > P(L/\vec{x})$, poruka $c$ se klasifikuje kao neželjena.
  - **Napomena**. Da bi smo mogli da klasifikujemo poruku neophodno je da odredimo $P(c)$ i $P(\vec{x}/c)$ za bilo koju poruku $m$ ali to očigledno ne može da se odredi precizno.
    - Međutim, možemo da aproksimiramo ove verovatnoće trening podacima.
    - Na primer verovatnoća $P(S)$ se može grubo odrediti izračunavanjem odnosa broja neželjenih poruka i broja svih poruka u trening podacima.

# Bajesovo filtriranje

- **Primena**.
  - Zbog jednostavnosti možemo smatrati da je karakterističan vektor binaran gde se prisustvo reči $w$ u poruci $m$ reprezentuje jednom jedinicom. Tada je:

$$P(x_w = 1/S) \approx \frac{broj\ neželjenih\ poruka\ u\ kojima\ je\ w\ obeleženo}{broj\ svih\ neželjenih\ poruka}$$

  - U suštini mi predstavljamo prisutnost reči $w_i$ u poruci $m$ vrednošću 1 u karakterističnom vektoru: $\vec{x} = (x_1, x_2, \dots, x_n)$.
  - Trening algoritam za Bajesov kalasifikator u tom slučaju mora da računa $2^n$ vrednosti $x$ što je nepraktično.

- **Primena**.
  - Kako bismo to izbegli uvodimo pretpostavku da su dve prisutne reči uslovno nezavisne jedna od druge što znači da je:

$$P(\vec{x}/c) = \prod_{i=1}^{n} P(\vec{x}_i/c)$$

$$\Lambda = \prod_{i=1}^{n} \Lambda_i(\vec{x}_i)$$

  - Pri čemu $\Lambda(\vec{x})$ predstavlja količnik $\frac{P(\vec{x}/S)}{P(\vec{x}/L)}$.

# Bajesovo filtriranje

- **Primena**.
  - Parametar $\lambda$ ukazuje na rizik kada klasifikujemo legitimnu poruku kao neželjenu:

$$\lambda = \frac{\mathcal{L}(L,S)}{\mathcal{L}(S,L)}$$

  - Funkcija $\mathcal{L}(c_1, c_2)$ određuje cenu loše klasifikacije pojavom klase $c_1$ umesto klase $c_2$.
  - Logično, $\mathcal{L}(S,S) = \mathcal{L}(L,L) = 0$.
  - Funkciju rizika definišemo na sledeći način:

$$R(c/\vec{x}) = \mathcal{L}(S,c)P(S/\vec{x}) + \mathcal{L}(L,c)P(L/\vec{x})$$

# Bajesovo filtriranje

- **Primena**.
  - Klasifikovanje legitimne poruke kao neželjene je generalno mnogo veća greška od dopuštanja da neželjena poruka prođe kroz filtar.
  - Ukoliko imamo dva tipa grešaka $L \rightarrow S$ i $S \rightarrow L$ , pretpostavljamo da je u slučaju greške tipa $L \rightarrow S$ parametar $\lambda$ nekoliko puta veći nego u slučaju greške $S \rightarrow L$.
  - Poruka se klasifikuje kao neželjena ukoliko je sledeći kriterijum ispunjen:

$$\frac{P(\vec{x}/c = S)}{P(\vec{x}/c = L)} > \lambda$$

# Bajesovo filtriranje

- **Trening algoritam**:
  1. **for all** $c \in \{S, L\}$ **do**
  2. *Izračunaj* $p(c)$
  3. **for all** $x_w \in \{0,1\}$ **do**
  4. *Izračunaj* $p(\vec{x}_w/c)$
  5. **end for**
  6. **end for**
  7. **for all** $\vec{x}_w \in \{0,1\}$ **do**
  8. *Izračunaj* $p(\vec{x}_w/c)$ koristeći Bajesovo pravilo
  9. **end for**
  10. **for all** $x_w \in \{0,1\}$ **do**
  11. *Izračunaj* $\Lambda(\vec{x}_w)$
  12. **end for**
  13. *Izračunaj* $\lambda \frac{P(L)}{P(S)}$

# Bajesovo filtriranje

- **Algoritam za klasifikaciju**:
  1. *Izračunati vektor* $\vec{x}_w$ *ulazne poruke* $m$.
  2. **if** $\Lambda(\vec{x}_w) > \lambda \frac{P(L)}{P(S)}$ **then**
  3. *poruka* $m$ je neželjena
  4. **else**
  5. poruka $m$ je legitimna
  6. **end if**

# Metoda najbližih suseda

- **Uvodne napomene**.
  - Ovaj metod se naziva još i metod zasnovan na memoriji ili instancama.
  - Pripada porodici algoritama učenja koji umesto izvođenja eksplicitne generalizacije upoređuje nove instance sa instancama koje se nakon trening faze nalaze u memoriji.
  - Učenje zasnovano na instancama je vrsta **lenjog učenja**.
    - Lenjo učenje u veštačkoj inteligenciji je poželjna osobina učenja ako postoji zahtev za stalnim menjanjem baze znanja, gde svaka takva promena ne povlači ponavljanje celog postupka učenja već samo efikasno inkrementalno dodavanje znanja.
  - Zajednička karakteristika ovih metoda je da svi oni smeštaju trening instance u **memorijske strukture** i koriste ih direkrno za klasifikaciju.
    - Najednostavniji oblik memorijske strukture je **višedimenzionalni prostor** definisan atributima u inicijalni vektor.
    - Svaka trening instanca je reprezentovana kao tačka u prostoru.
  - Procedura klasifikacije je najčešće jednostavna varijanta algoritma **k najbližih suseda**.

# Metoda najbližih suseda

- **Klasifikacija**.
  - Klasifikacija novih instanci se obavlja prema principu najbližeg suseda, gde se nova instanca upoređuje sa instancama iz skupa za učenje korišćenjem definisane metrike.
  - Metrika definiše rastojanje instanci na osnovu vrednosti njihovih atributa, a odgovara intuitivnom shvatanju sličnosti instanci tako da ako su instance sličnije rastojanje je manje.
  - Nova instanca se klasifikuje na osnovu pretraživanja skupa za učenje sa ciljem pronalaženja instance koja mu je u smislu rastojanja najbliža.

# Metoda najbližih suseda

- **Ideja algoritma k najbližih suseda**.
  - **Trening**.
    - Smeštanje trening poruka.
  - **Klasifikacija**.
    - Konkretna poruka $x$ sadrži $k$ najbližih suseda među porukama u trening skupu. Ukoliko ima više spam poruka među ovim susedima, poruka će biti klasifikovana kao spam, u suptrotnom klasifikator će je klasifikovati kao legitimnu poruku.
  - Kao što se može primetiti, trening faza ne postoji u običajenom smislu.
  - Cena ovoga je sporija procedura odlučivanja jer u cilju klasifikovanja jedne poruke moramo računati rastojanja do svih trening poruka i pronaći $k$ najbližih suseda.

# Metoda najbližih suseda

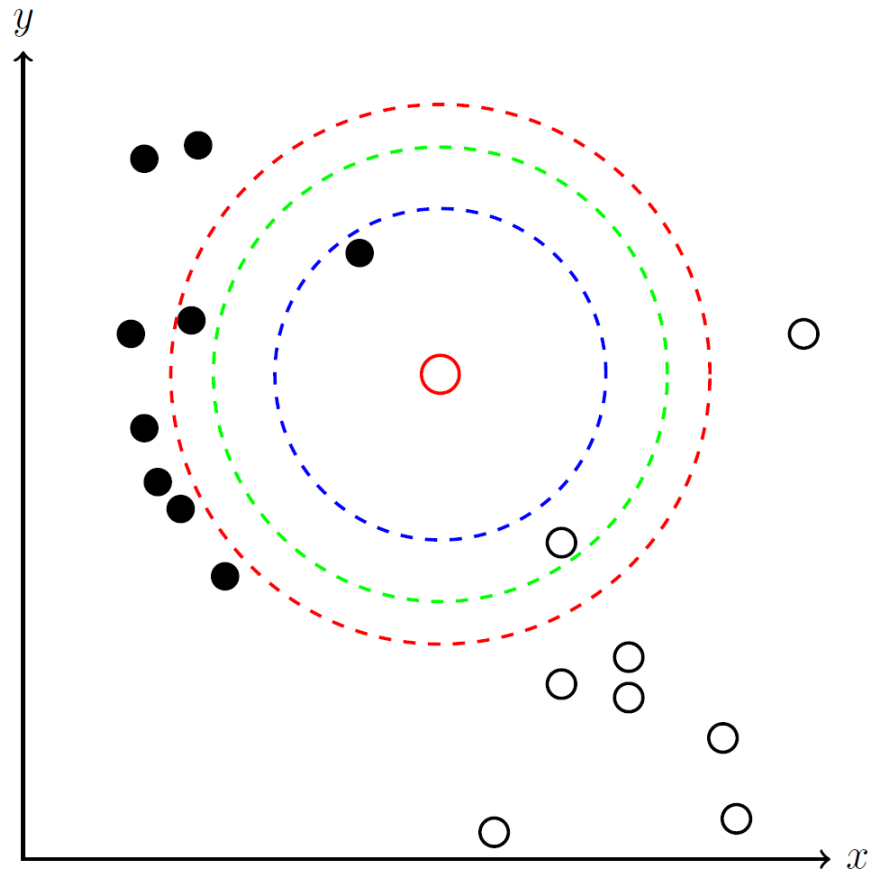
- **Koji je problem karakterističan za ovaj algoritam**?
  - Mana klasifikatora zasnovanih na memoriji je **cena izračunavanja u fazi klasifikacije** zbog njihove "lenje" prirode.
    - Svi trening primeri moraju da sprovedu klasifikaciju, što je sa velikim skupovima podataka posebno teško.
  - Drugi problem koji se javlja kod ovog algoritma je nepostojanje parametara čije će podešavanje smanjiti broj lažno pozitivnih instanci.
    - Taj problem se jednostavno rešava modifikacijom klasifikacionog pravila na sledeće $l/k$ pravilo:
      - Ukoliko je $l$ ili više poruka među $k$ najbližih suseda poruke $x$ *spam*, klasifikuj $x$ kao spam, u suprotnom klasifikuj $x$ kao legitimnu poruku.

# Metoda najbližih suseda

- **Pseudo-kod algoritma**:
  1. **Ulaz**: *k* – broj suseda, *z* – nova instanca opisana sa *m* obeležja, *TS* – trening skup od *n* instanci opisan sa *m* obeležja, *Kat* – klase instanci iz trening skupa
  2. **Izlaz**: *c* – predviđena kategorija nove instance
  3. **procedure** kNAJBLIŽIHSUSEDA (*k*, *z*, *TS*, *Kat*)
  4. **for all** $x \in \{TS\}$ do
     // računanje rastojanja od *z* do svih ostalih instanci *x* iz trening skupa
  5. *d*(*x*) = *rastojanje*(*x*,*z*);
  6. **end for**
  7. [*sortirane_instance*; *index*] = *sort*(*d*); // sortiranje distanci u rastućem poretku
  8. *kategorija_suseda* = *Kat*(*index*(1 : *k*));
  9. *c* = *vecinaGlasova*(*kategorija_suseda*);
  10. **return** *c*
  11. **end procedure**

# Metoda najbližih suseda

- **Primer**. Ilustracija algoritma za *k*=1, *k*=2 i *k*=3.




* Slika preuzeta iz [2].

# Metoda najbližih suseda

- **Pretraga prostora rešenja**.
  - Neke varijante klasifikacije zasnovane na instancama nastoje da redukuju broj instance u skupu za učenje, prvenstveno radi smanjenja opsega pretraživanja pri klasifikaciji novih instanci, jer skup instanci za učenje po pravilu sadrži veliki broj redundantnih instanci.
  - Kod problema klasifikacije najvažnije su instance koje se nalaze u blizini granica među klasama, a instance iz unutrašnjosti određenog područija klasa mogu se izostaviti bez posledica na tačnost klasifikacije.

# Metoda najbližih suseda

- **Pretraga prostora rešenja**.
  - Kriterijumi prihvatanja i eleminisanja instanci uglavnom pretstavljaju nepovratne strategije pohlepnog karaktera.
  - Najjednostavniji kriterijum je da se ispituje instanca prema rezultatu klasifikacije korišćenjem do tada **izdvojenih reprezentativnih instanci**.
    - Ako je u pitanju **netačna klasifikacija** instance ona se pridodaje u skup reprezentativnih instanci jer je evidentno da menja granice klasa.
    - Ukoliko je **klasifikacija instance tačna** tada se ona proglašava suvišnom jer informacija koju nosi je već sadržana u skupu pomoću kojeg je klasifikovana.
  - Nedostaci ovog kriterijuma za izbor instanci:
    - U početnoj fazi procesa pretraživanja postoji nezanemariva verovatnoća odbacivanja instanci koje se mogu pokazati važnim za tačnost klasifikacije rezultujućeg modela.
    - Izabrani podskup reprezentativnih instanci ne zavisi samo od polaznog skupa, već i od redosleda evaluacije instanci.

# Metoda najbližih suseda

- **Modfikovano Euklidsko rastojanje**.
  - Pored izbora instanci za pamćenje, na oblik klasifikacionog modela se može uticati i **modifikacijom funkcije rastojanja**.
  - Jednak uticaj svih atributa u instanci na konačan rezultat je jedno od svojstava euklidskog rastojanja, ali su u praksi retki problemi kod kojih svi atributi imaju jednaku vrednost za proces klasifikacije.
  - Modifikacija euklidskog rastojanja podrazumeva uvođenje **težinskih vrednosti atributa**.
  - Ako sa $w_i$ označimo težinsku vrednost pridruženu atributu $A_i$, onda modifikovano Euklidsko rastojanje instanci $x$ i $y$ možemo predstaviti na sledeći način:

$$d_w(x, y) = \sqrt{\sum_{i=1}^{n} w_i^2 (x_i - y_i)^2}$$

# Obrada korpusa

- **Jezička obrada kolekcije dokumenata**.
  - Osnovne faze jezičke obrade teksta su:
    - tokenizacija,
    - izbacivanje čestih reči,
    - normalizacija,
    - lematizacija.

# Obrada korpusa

- **Jezička obrada kolekcije dokumenata**.
  - Tokenizacija podrazumeva deljenje teksta na tokene, tj. nizove znakova koji predstavljaju semantičke jedinice pogodne za dalju obradu.
  - Intuitivno, tokenizacija se može posmatrati kao deljenje teksta na pojedine reči, pri čemu se zanemaruju znakovi interpunkcije.
  - Primer:
    - “Данас стижу резултати анализе млека и производа од млека”
  - Rezultat tokenizacije je sledeći niz reči:
    - Данас стижу резултати анализе млека и производа од млека

# Obrada korpusa

- **Jezička obrada kolekcije dokumenata**.
  - Za prethodni primer, pretpostavljamo da su česte reči (tzv. “stop-reči”):
    - “од”
    - “и”
  - Pretpostavlja se da stop reči nose malu informacionu vrednost za klasifikaciju.
  - Rezulat tokenizacije:
    - Данас стижу резултати анализе млека и производа од млека
  - Nakon uklanjanja stop-reči dobijamo:
    - Данас стижу резултати анализе млека производа млека

# Obrada korpusa

- **Jezička obrada kolekcije dokumenata**.
  - **Normalizacija** podrazumeva proces izjednačavanja tokena koji se samo površinski razlikuju.
  - Uklanjanje **površinskih razlika** može da uključuje:
    - prevođenje tokena na isto pismo,
    - korišćenje samo malih slova za sve tokene,
    - ujednačavanje slovnih skraćenica, itd.
  - Reultat nakon izbacivanja stop-reči u prethodnom primeru bio je:
    - Данас стижу резултати анализе млека производа млека
  - Rezultat normalizacije koja obuhvata i prevođenje tokena na latinično pismo je:
    - danas stižu rezultati analize mleka proizvoda mleka
  - Napomena: normalizacija može da podraumeva i uklanjanje dijakritičkih znakova, iako njihovo uklanjanje može da promeni značenje reči.
    - Primer – uporedite značenje reči “šišanje” i “sisanje”.

# Obrada korpusa

- **Jezička obrada kolekcije dokumenata**.
  - Neki tokeni se ne razlikuju samo površinski, već predstavljaju različite pojavne oblike iste reči.
    - Primer: tokeni “čoveku” i “ljudi” predstavljaju oblike imenice “čovek” koje želimo prilikom klasifikacije da tretiramo na isti način.
  - Zbog toga jezička obrada teksta uključuje i fazu **lematizacije**, koja podrazumeva svođenje izvedenih oblika reči na osnovne oblike (leme).
  - U postupku lematizacije:
    - imenice se svode na nominativ jednine: (“softveri” $\rightarrow$ “softver”, “psa” $\rightarrow$ “pas”),
    - glagoli na infinitiv (“išli” $\rightarrow$ “ići”, “tučem” $\rightarrow$ “tući”), itd.
  - Rezultat normalizacije:
    - danas stižu rezultati analize mleka proizvoda mleka
  - Nakon lematizacije dobijamo sledeće termine:
    - danas stići rezultat analiza mleko proizvod mleko

# Obrada korpusa

- **Izdvanjanje obeležja**.
  - BOW model (*bag of words*).
    - Za dati set termina (tokena) $T = \{t_1, t_2, \dots, t_N\}$ dokument $d$ se predstavlja kao N-dimenzioni vector obeležja $x = \{x_1, x_2, \dots, x_N\}$, gde $x_i$ zavisi od pojavljivanja termina $t_i$ u dokumentu $d$.
    - Problem: odabrati *N* najznačajnijih obeležja kako bi se smanjila dimenzionalnost.
  - "*Bag of character n-grams*" model.
  - *Sparse Binary Polynomial Hashing*.
  - *Orthogonal Sparse Bigrams*.
  - …

# Obrada korpusa

- **Izbor obeležja**.
  - Cilj:
    - povećati tačnost klasifikatora i
    - povećati brzinu rada klasifikatora (sprečite prokletstvo visoke dimenzionalnosti).
  - Metode:
    - Informaciona dobit (*information gain*, IG)
    - …

# Izabrani napadi na filtre neželjene elektronske pošte

- **Maskiranje reči**.
  - Reči poput “sex”, “free”, “congratulations” imaju veliki značaj za klasifikaciju poruke kao neželjene.
  - Tipični vidovi maskiranja:
    - f-r-e-e (dodati specijalni karakteri)
    - fr<!--xx-->ee
    - \item <a href=’mailto:%66ree’>free</a>
    - \item o fr&#101xe

# Izabrani napadi na filtre neželjene elektronske pošte

- **Trovanje Bajesovog klasifikatora**.
  - Statistički filtri su ranjivi na napade koji se izvode dodavanjem slučajnih ili pažljivo odabranih naizgled legitimnih reči – filtar neželjene pošte može ovakvu neželjenu poruku da klasifikuje kao legitimnu (**statistička greška II tipa**).
    - Reči se mogu preuzeti sa izvora poput Reuters članaka, pisanih tekstova na engleskom jeziku itd.
  - Ovakvi napadi često izazivaju **statističku grešku I tipa**.
    - Statistička greška I tipa se odnosi na povećanu učestalost lažno pozitivnih alarma.
    - Razlog pojave ove greške leži u tome što korisnik obučava klasifikator zatrovanim podacima.

# Izabrani napadi na filtre neželjene elektronske pošte

- **Ostali značajniji napadi**. *
  - “Backscatter spam”.
  - “Image spam”.
  - “Botnet spam”.

* Više detalja o ovim napadima možete naći u [1].

# Pecanje


** Slika preuzete sa lokacije: https://www.tripwire.com/state-of-security/security-awareness/6-common-phishing-attacks-and-how-to-protect-against-them/*

# Pecanje

- **Pecanje** (engl. *phishing*) predstavlja vrstu kriminalne aktivnosti koja koristi tehnike društvenog inženjeringa (prevara) pomoću koje napadači dolaze do osetljivih informacija kao što su:
  - lozinke,
  - detalji o kreditnim karticama, itd.
- Pecanje se najčešće izvodi **pomoću e-pošte**.
  - Poruke izgledaju kao da su ih poslala lica ili institucije od poverenja.
  - Poruke izgledaju kao zvanična elektronska komunikacija.
  - Cilj: navođenje korisnika da na lažnoj Veb lokaciji banke ili druge finansijske insitutcije ostavi podatke koji ne smeju da budu dostupni drugima.
- Na taj način napadač dolazi do podataka kao što su brojevi kreditnih kartica, računa, korisnička imena i lozinke za pristup pravim računima.

# Krađa identiteta

- Krađa identiteta je vrsta prevare kojom se od korisnika putem lažne poruke e-pošte ili web lokacije prikupljaju lični i finansijski podaci.
- Tipičan scenario (neželjena e-pošta + pecanje):
  - Žrtvi se šalje se e-pošta koja je slična službenom obaveštenju (npr. iz banke ili druge ustanove).
  - Žrtva se upućuje na lažnu Veb lokaciju.
  - Na lokaciji se od žrtve traži da unese lične podatke, poput broj računa, broja kreditne kartice ili lozinke.
  - Unešeni podaci se dalje koriste za krađu identiteta.

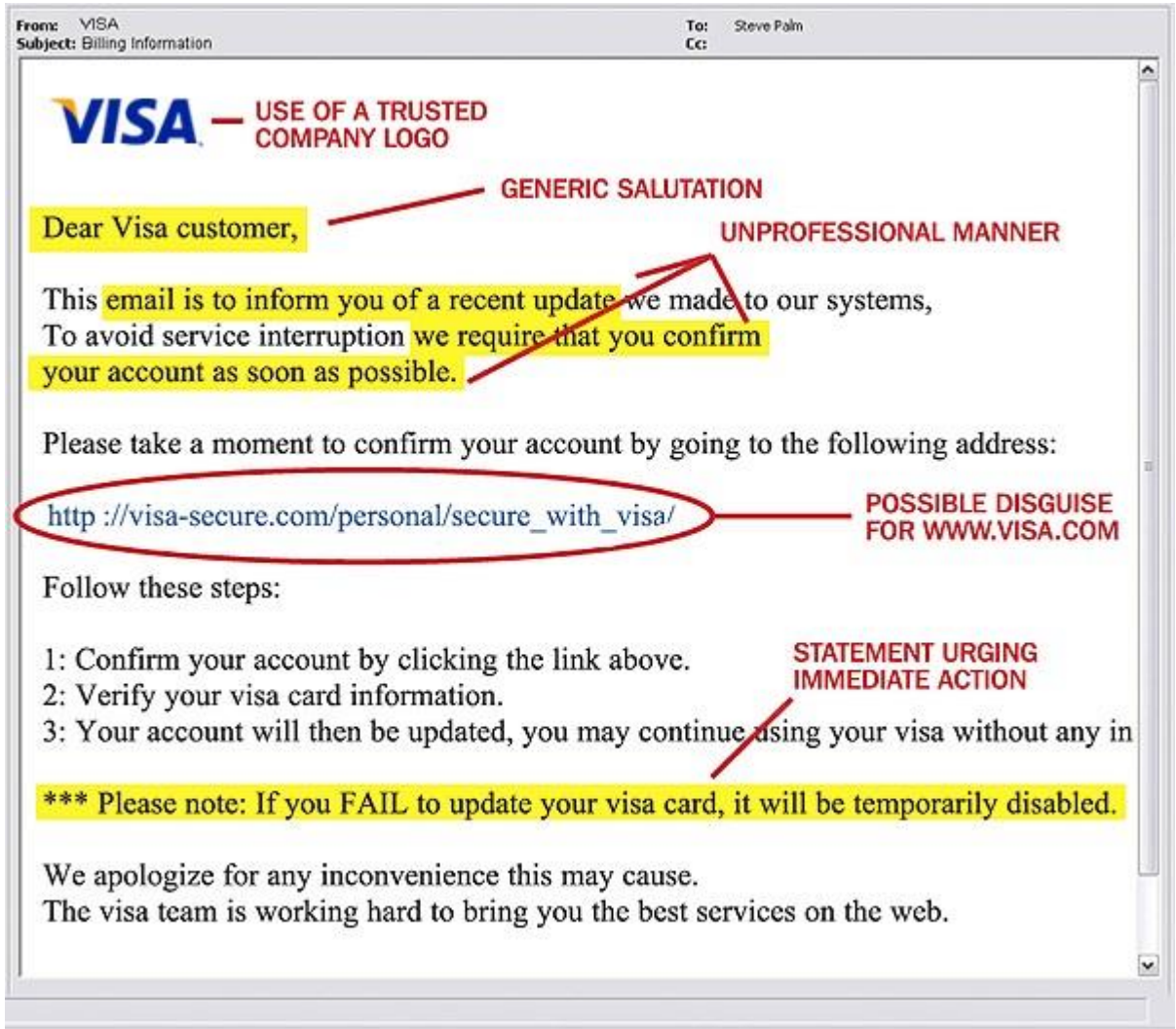
# Krađa identiteta

- Moguće posledice:
  - podnošenje zahteva za kredit,
  - prebacivanje novca sa jednog računa na drugi i naknadno poništenje računa.
- U ovom slučaju Vi postajete svesni da vam je identitet preuzet kada primetite da novca nema!

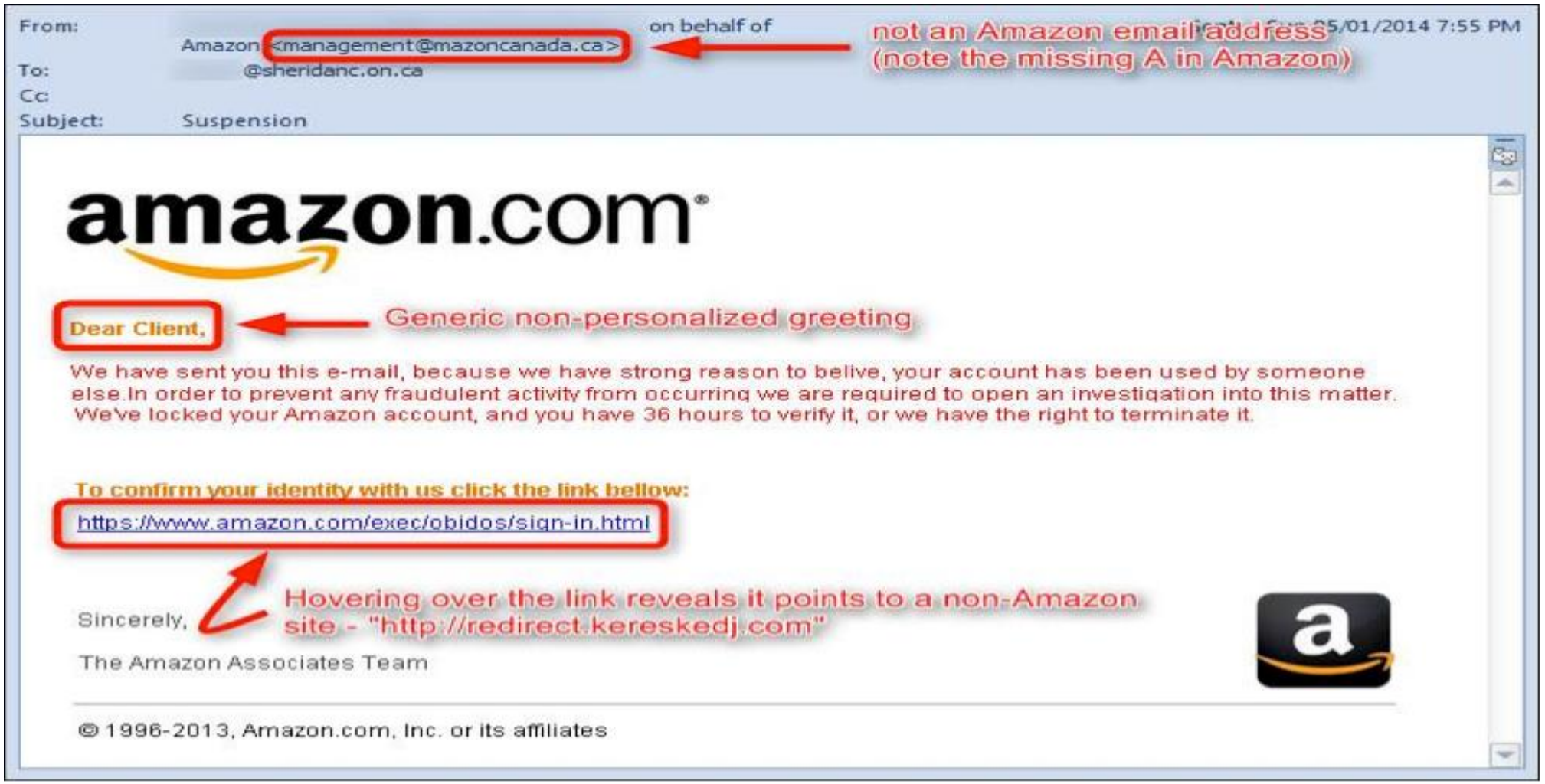
# Krađa identiteta

- Kako da se zaštitite? Uočite prevaru pre nego što bude kasno!
  - Poslovna politika: **lični podaci se ne traže putem e-pošte**.
    - Ako dobijete ovakav zahtev, priupitajte se da li je zahtev legitiman.
    - Još bolje, pozovite instituciju i pitajte.
    - Dodatno proverite da li se odgovor traži **u vrlo kratkom roku**, kako bi se izbegla moguća šteta.
    - Ovakvi zahtevi najčešće nisu personalizovni.
  - Primer:
    - Poštovani (**nema imena**), neophodno je da ažurirate informacije u vašem nalogu s ciljem dobijanja prijava o neaktivnosti, prevarama i lažnom predstavljanju. Izbegavanje ažuriranja vaših zapisa dovešće do brisanja naloga (**šteta**). Da biste potvrdili vaše podatke, molimo vas sledeće podatke ...

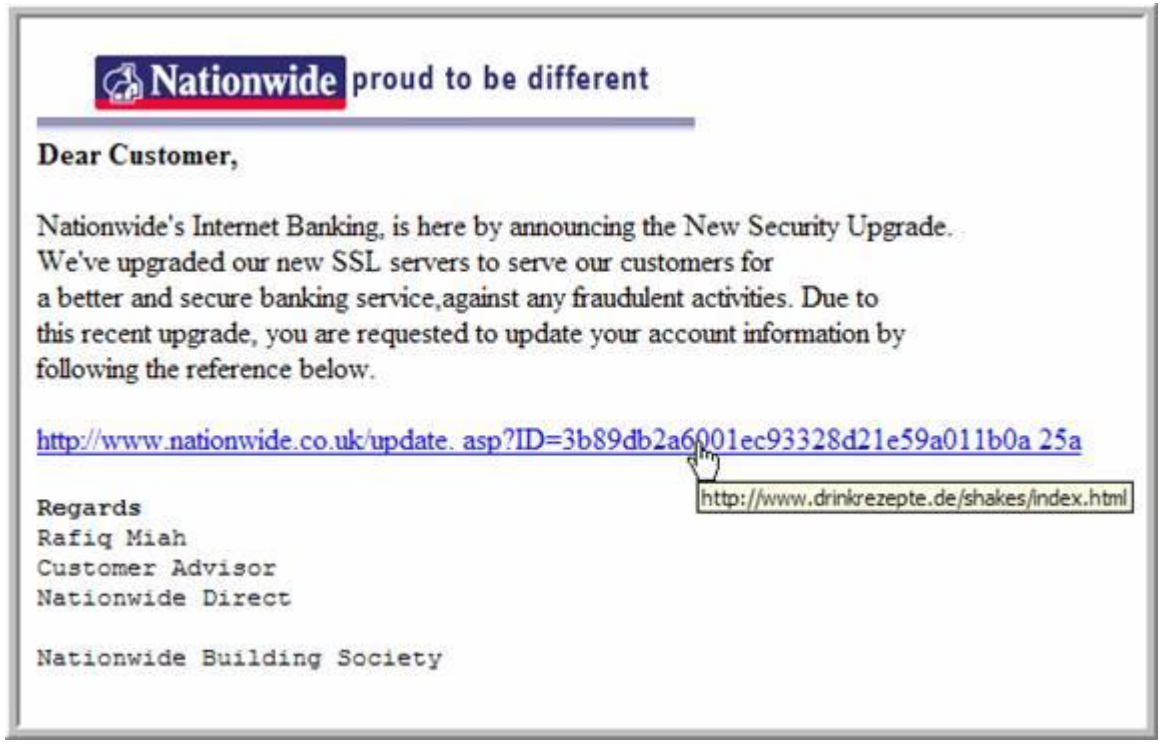
# Krađa identiteta



# Krađa identiteta



# Krađa identiteta



Nationwide proud to be different

**Dear Customer,**

Nationwide's Internet Banking, is here by announcing the New Security Upgrade. We've upgraded our new SSL servers to serve our customers for a better and secure banking service,against any fraudulent activities. Due to this recent upgrade, you are requested to update your account information by following the reference below.

http://www.nationwide.co.uk/update. asp?ID=3b89db2a6001ec93328d21e59a011b0a 25a

http://www.drinkrezepte.de/shakes/index.html

Regards
Rafiq Miah
Customer Advisor
Nationwide Direct

Nationwide Building Society

# Literatura


1. A. Bhowmick, S. M. Hazarika (2016): Machine Learning for E-mail Spam Filtering – Review, Techniques and Trends. https://arxiv.org/pdf/1606.01042.pdf
2. A. Nedeljković (2015): Implementacija i evaluacija algoritama mašinskog učenja za filtriranje neželjene elektronske pošte. Matematički fakultet, Univerzitet u Beogradu.
3. M. Gnjatović (2017): Uvod u pronalaženje informacija na Vebu. Visoka škola elektrotehnike i računarstva strukovnih studija, Beograd. http://www.gnjatovic.info/pronalazenjeinformacija/index.html

# Hvala na pažnji

**Pitanja su dobrodošla.**