

```
In [1]: import pandas as pd
import re
import matplotlib.pyplot as plt
%matplotlib inline
```

Ambil data dari CSV

```
In [2]: import pandas as pd
data = pd.read_csv("javaheritagehotel.csv", delimiter=";") # read file csv nya dulu y
data
```

Out[2]:

	customer_name	customer_judul_komen	customer_judul_komen_url	customer
0	andy S	Hotel yang nyaman	https://www.tripadvisor.co.id/ShowUserReviews-...	Salah satu Pun
1	Fabian Hernanda	Keluarga Hernawan	https://www.tripadvisor.co.id/ShowUserReviews-...	Mobil I Ja Kelu
2	christiandi a	Kamarnya nyaman,kolam renangnya aman	https://www.tripadvisor.co.id/ShowUserReviews-...	Hotel i pilihan p
3	Fajar Heks	Hotel juara	https://www.tripadvisor.co.id/ShowUserReviews-...	Hi purwoke
4	Yono Suyono	nyaman dan tenang	https://www.tripadvisor.co.id/ShowUserReviews-...	bai fasilitas
...
114	Irwan_Joe	Dari dinasti ke horison	https://www.tripadvisor.co.id/ShowUserReviews-...	Ini adal kalinya :
115	Rachel U	Semoga dapat menjadi lebih baik	https://www.tripadvisor.co.id/ShowUserReviews-...	Pertama bangun
116	Anu_Singapore	Di atas rata-rata	https://www.tripadvisor.co.id/ShowUserReviews-...	Kamar hotel ya
117	agus budiman	harga rata-rata	https://www.tripadvisor.co.id/ShowUserReviews-...	sala purwoke
118	pleasantinterlude	Baik tetapi dapat ditingkatkan	https://www.tripadvisor.co.id/ShowUserReviews-...	Ting untuk m p

119 rows × 4 columns

In [3]:

```
pip install Sastrawi
```

Requirement already satisfied: Sastrawi in c:\users\user\anaconda3\lib\site-packages (1.0.1)

Note: you may need to restart the kernel to use updated packages.

Pembersihan Data

In [4]:

```
import Sastrawi
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

slangs={
    "@": "di", "abis": "habis", "ad": "ada", "adlh": "adalah", "afaik": "as far",
    "ahaha": "haha", "aj": "saja", "ajep-ajep": "dunia gemerlap", "ak": "saya",
    "akkoh": "aku", "akuwh": "aku", "alay": "norak", "alow": "halo", "ambilin":
    "ancur": "hancur", "anjrit": "anjing", "anter": "antar", "ap2": "apa-apa", "
    "apes": "sial", "aps": "apa", "aq": "saya", "aquwh": "aku", "asbun": "asal b
    "asekk": "asyik", "asem": "asam", "aspal": "asli tetapi palsu", "astul": "as
    "au ah": "tidak mau tahu", "awak": "saya", "ay": "sayang", "ayank": "sayang"
    "bakalan": "akan", "bandes": "bantuan desa", "bangedh": "banget", "banpol":
    "banpur": "bantuan tempur", "basbang": "basi", "bcanda": "bercanda", "bdg":
    "begajulan": "nakal", "beliin": "belikan", "bencong": "banci", "bentar": "se
    "beresin": "membereskan", "bete": "bosan", "beud": "banget", "bg": "abang",
    "bgt": "banget", "bijimane": "bagaimana", "bintal": "bimbingan mental", "bkl
    "bknya": "bukannya", "blegug": "bodoh", "blh": "boleh", "bln": "bulan", "bl
    "bnran": "yang benar", "bodor": "lucu", "bokap": "ayah", "boker": "uang air
    "boljug": "boleh juga", "bonek": "bocah nekat", "boyeh": "boleh", "br": "bar
    "bro": "saudara laki-laki", "bru": "baru", "bs": "bisa", "bsen": "bosan", "b
    "buaya": "tidak setia", "bubbu": "tidur", "bubu": "tidur", "bumil": "ibu ham
    "bwt": "buat", "byk": "banyak", "byrin": "bayarkan", "cabal": "sabar", "cada
    "calo": "makelar", "can": "belum", "capcus": "pergi", "caper": "cari perhati
    "cekal": "cegah tangkal", "cemen": "penakut", "cengengesan": "tertawa", "cep
    "cew": "cewek", "chuyunk": "sayang", "cimeng": "ganja", "cipika cipiki": "ci
    "ciyh": "sih", "ckepp": "cakep", "ckp": "cakep", "cmiiw": "correct me if i'm
    "cong": "banci", "conlok": "cinta lokasi", "cowwyy": "maaf", "cp": "siapa",
    "cucok": "cocok", "cuex": "cuek", "cumi": "Cuma miscall", "cups": "culun",
    "curanmor": "pencurian kendaraan bermotor", "curcol": "curahan hati colongan
    "cwek": "cewek", "cyin": "cinta", "d": "di", "dah": "deh", "dapet": "dapat",
    "dek": "adik", "demen": "suka", "deyh": "deh", "dgn": "dengan", "diancurin":
    "dimaafin": "dimaafkan", "dimintak": "diminta", "disono": "di sana", "dket":
    "dkk": "dan kawan-kawan", "dll": "dan lain-lain", "dlu": "dulu", "dngn": "de
    "doku": "uang", "dongs": "dong", "dpt": "dapat", "dri": "dari", "drmn": "dar
    "dst": "dan seterusnya", "dtg": "datang", "duh": "aduh", "duren": "durian",
    "egp": "emang gue pikirin", "eke": "aku", "elu": "kamu", "emangnya": "memang
    "endak": "tidak", "enggak": "tidak", "envy": "iri", "ex": "mantan", "fax": "
    "fifo": "first in first out", "folbek": "follow back", "fyi": "sebagai infor
    "gag": "tidak", "gaje": "tidak jelas", "gak papa": "tidak apa-apa", "gan": "
    "gatek": "gagap teknologi", "gawe": "kerja", "gbs": "tidak bisa", "gebetan":
    "geje": "tidak jelas", "gepeng": "gelandangan dan pengemis", "ghiy": "lagi",
    "gimana": "bagaimana", "gino": "gigi nongol", "github": "gitu", "gj": "tidak
    "gn": "begini", "goblok": "bodoh", "golput": "golongan putih", "gowes": "men
    "gpony": "tidak punya", "gr": "gede rasa", "gretongan": "gratisan", "gtau": "
    "gua": "saya", "guoblok": "goblok", "gw": "saya", "ha": "tertawa", "haha": "
    "hallow": "halo", "hankam": "pertahanan dan keamanan", "hehe": "he", "helo":
    "hlm": "halaman", "hny": "hanya", "hoax": "isu bohong", "hr": "hari", "hrus":
    "hubdar": "perhubungan darat", "huff": "mengeluh", "hum": "rumah", "humz": "
    "ilang": "hilang", "ilfil": "tidak suka", "imho": "in my humble opinion", "i
    "item": "hitam", "itungan": "hitungan", "iye": "iya", "ja": "saja", "jadiin":
    "jaim": "jaga image", "jayus": "tidak lucu", "jdi": "jadi", "jem": "jam", "j
    "jgnkan": "jangan", "jir": "anjing", "jln": "jalan", "jomblo": "tidak pun
```

"jubil": "juru bicara", "jutek": "galak", "k": "ke", "kab": "kabupaten", "ka
 "kacrut": "kacau", "kativ": "kepala divisi", "kagak": "tidak", "kalo": "kala
 "kamtibmas": "keamanan dan ketertiban masyarakat", "kamuwh": "kamu", "kanwil
 "karna": "karena", "kasubbag": "kepala subbagian", "katrok": "kampungan", "k
 "kbr": "kabar", "kdu": "harus", "kec": "kecamatan", "kejurnas": "kejuaraan n
 "kekeuh": "keras kepala", "kel": "kelurahan", "kemaren": "kemarin", "kepenge
 "kepsek": "kepala sekolah", "kesbang": "kesatuan bangsa", "kesra": "kesejaht
 "ketrima": "diterima", "kgiatan": "kegiatan", "kibul": "bohong", "kimpoi": "
 "klian": "kalian", "kloter": "kelompok terbang", "klw": "kalau", "km": "kam
 "kmrn": "kemarin", "knal": "kenal", "knp": "kenapa", "kodya": "kota madya",
 "komdis": "komisi disiplin", "komsov": "komunis soviet", "kongkow": "kumpul
 "kopdar": "kopi darat", "korup": "korupsi", "kpn": "kapan", "krenz": "keren"
 "kyk": "seperti", "la": "lah", "lam": "salam", "lamp": "lampiran", "lanud": "
 "latgab": "latihan gabungan", "lebay": "berlebihan", "leh": "boleh", "lelet"
 "lemot": "lambat", "lgi": "lagi", "lgsg": "langsung", "liat": "lihat", "litb
 "lmyn": "lumayan", "lo": "kamu", "loe": "kamu", "lola": "lambat berfikir", "
 "lp": "lupa", "luber": "langsung, umum, bebas, dan rahasia", "luchuw": "lucu
 "lwn": "lawan", "maacih": "terima kasih", "mabal": "bolos", "macem": "macam"
 "maem": "makan", "magabut": "makan gaji buta", "maho": "homo", "mak jang": "
 "malem": "malam", "mam": "makan", "maneh": "kamu", "maniez": "manis", "mao":
 "melu": "ikut", "mepet": "dekat sekali", "mgu": "minggu", "migas": "minyak d
 "mikol": "minuman beralkohol", "miras": "minuman keras", "mlah": "malah", "m
 "mo": "mau", "mokad": "mati", "moso": "masa", "mpe": "sampai", "msk": "masuk
 "mt": "makan teman", "mubes": "musyawarah besar", "mulu": "melulu", "mumpung
 "munas": "musyawarah nasional", "muntaber": "muntah dan berak", "musti": "me
 "mw": "now watching", "n": "dan", "nanam": "menanam", "nanya": "bertanya", "
 "napi": "narapidana", "napza": "narkotika, alkohol, psikotropika, dan zat ad
 "narkoba": "narkotika, psikotropika, dan obat terlarang", "nasgor": "nasi go
 "ndiri": "sendiri", "ne": "ini", "nekolin": "neokolonialisme", "nembak": "me
 "ngabuburit": "menunggu berbuka puasa", "ngaku": "mengaku", "ngambil": "meng
 "nganggur": "tidak punya pekerjaan", "ngapah": "kenapa", "ngaret": "terlamba
 "ngebandel": "berbuat bandel", "ngegosip": "bergosip", "ngeklaim": "mengklai
 "ngeles": "berkilah", "ngelidur": "menggigau", "ngerampok": "merampok", "ngg
 "ngiler": "mau", "ngiri": "iri", "ngisiin": "mengisikan", "ngmng": "bicara",
 "ngubek2": "mencari-cari", "ngurus": "mengurus", "nie": "ini", "nih": "ini",
 "nntn": "nonton", "nobar": "nonton bareng", "np": "now playing", "ntar": "na
 "numpuk": "bertumpuk", "nutupin": "menutupi", "nyari": "mencari", "nyekar":
 "nyoblos": "mencoblos", "nyokap": "ibu", "ogah": "tidak mau", "ol": "online"
 "oot": "out of topic", "org2": "orang-orang", "ortu": "orang tua", "otda": "
 "otw": "on the way, sedang di jalan", "pacal": "pacar", "pake": "pakai", "pa
 "pansus": "panitia khusus", "parpol": "partai politik", "pasutri": "pasangan
 "pede": "percaya diri", "pelatnas": "pemusatan latihan nasional", "pemda": "
 "pemekot": "pemerintah kota", "pemred": "pimpinan redaksi", "penjas": "pendid
 "perda": "peraturan daerah", "perhatiin": "perhatikan", "pesenan": "pesanan"
 "pilkada": "pemilihan kepala daerah", "pisan": "sangat", "pk": "penjahat kel
 "pmrnth": "pemerintah", "polantas": "polisi lalu lintas", "ponpes": "pondok
 "prg": "pergi", "prnh": "pernah", "psen": "pesan", "pst": "pasti", "pswt": "
 "qmu": "kamu", "rakor": "rapat koordinasi", "ranmor": "kendaraan bermotor",
 "rehab": "rehabilitasi", "rempong": "sulit", "repp": "balas", "restik": "res
 "rmh": "rumah", "ru": "baru", "ruko": "rumah toko", "rusunawa": "rumah susun
 "saia": "saya", "salting": "salah tingkah", "sampe": "sampai", "samsek": "sa
 "satpam": "satuan pengamanan", "sbb": "sebagai berikut", "sbh": "sebuah", "s
 "scr": "secara", "sdgkn": "sedangkan", "sdt": "sedikit", "se7": "setuju", "
 "sembako": "sembilan bahan pokok", "sempet": "sempat", "sendratari": "seni d
 "shg": "sehingga", "siech": "sih", "sikon": "situasi dan kondisi", "sinetron
 "siramin": "siramkan", "sj": "saja", "skalian": "sekalian", "sklh": "sekolah
 "slesai": "selesai", "sll": "selalu", "slma": "selama", "slsai": "selesai",
 "sndiri": "sendiri", "soljum": "sholat jumat", "songong": "sombong", "sory":
 "sotoy": "sok tahu", "spa": "siapa", "sppa": "siapa", "spt": "seperti", "srt
 "stiap": "setiap", "stlh": "setelah", "suk": "masuk", "sumpek": "sempit", "s
 "tajir": "kaya", "tau": "tahu", "taw": "tahu", "td": "tadi", "tdk": "tidak",
 "telat": "terlambat", "telmi": "telat berpikir", "temen": "teman", "tengil":
 "tggu": "tunggu", "tgu": "tunggu", "thankz": "terima kasih", "thn": "tahun",
 "tipiwan": "TvOne", "tk": "terima kasih", "tlp": "telepon", "tls": "tulisi",

"tmen2": "teman-teman", "tmpah": "tumpah", "tmpt": "tempat", "tngu": "tunggu",
 "tokai": "tai", "toserba": "toko serba ada", "tpi": "tapi", "trdhulu": "terd",
 "trm": "terima", "trs": "terus", "trutama": "terutama", "ts": "penulis", "ts",
 "ttg": "tentang", "tuch": "tuh", "tuir": "tua", "tw": "tahu", "u": "kamu", "
 "ujg": "ujung", "ul": "ulangan", "unyu": "lucu", "uplot": "unggah", "urang": "
 "utk": "untuk", "valas": "valuta asing", "w/": "dengan", "wadir": "wakil dir",
 "warkop": "warung kopi", "warteg": "warung tegal", "wat": "buat", "wkt": "wa",
 "xixixi": "tertawa", "ya": "iya", "yap": "iya", "yaudah": "ya sudah", "yawda",
 "yl": "yang lain", "yo": "iya", "yowes": "ya sudah", "yup": "iya", "7an": "t",
 "acc": "accord", "adlah": "adalah", "adoh": "aduh", "aha": "tertawa", "aing",
 "ajj": "saja", "aka": "dikenal juga sebagai", "akko": "aku", "akku": "aku",
 "aljasa": "asal jadi saja", "ama": "sama", "ambl": "ambil", "anjir": "anjing",
 "apaan": "apa", "ape": "apa", "aplot": "unggah", "apva": "apa", "aqu": "aku",
 "aseek": "asyik", "asek": "asyik", "aseknya": "asyiknya", "asoy": "asyik", "
 "ath": "kalau begitu", "atuh": "kalau begitu", "ava": "avatar", "aws": "awas",
 "ayok": "ayo", "bacot": "banyak bicara", "bales": "balas", "bangdes": "pemba",
 "banpres": "bantuan presiden", "bansarkas": "bantuan sarana kesehatan",
 "basis": "badan amal, zakat, infak, dan sedekah", "bcoz": "karena", "beb": "
 "belum": "belum", "bener": "benar", "ber2": "berdua", "berdikari": "berdiri",
 "bet": "banget", "beti": "beda tipis", "beut": "banget", "bgd": "banget", "b",
 "bhuhu": "tidur", "bimbuluh": "bimbingan dan penyuluhan", "bisi": "kalau-kal",
 "blg": "bilang", "blm": "belum", "bls": "balas", "bnchi": "benci", "bngung": "
 "bohay": "badan aduhai", "bokep": "porno", "bokin": "pacar", "bole": "boleh",
 "bonyok": "ayah ibu", "bpk": "bapak", "brb": "segera kembali", "brngkt": "be",
 "brur": "saudara laki-laki", "bsa": "bisa", "bsk": "besok", "bu_bu": "tidur",
 "buber": "buka bersama", "bujubune": "luar biasa", "buser": "buru sergap", "
 "byr": "bayar", "c8": "chat", "cabut": "pergi", "caem": "cakep", "cama-cama",
 "cangcut": "celana dalam", "cape": "capek", "caur": "jelek", "cekak": "tidak",
 "cemplungin": "cemplungkan", "ceper": "pendek", "ceu": "kakak perempuan", "c",
 "cin": "cinta", "ciye": "cie", "ckck": "ck", "clbk": "cinta lama bersemi kem",
 "cnenk": "senang", "congor": "mulut", "cow": "cowok", "coz": "karena", "cpa",
 "gombal": "suka merayu", "gpl": "tidak pakai lama", "gpp": "tidak apa-apa",
 "gtw": "tidak tahu", "gue": "saya", "guys": "teman-teman", "gws": "cepat sem",
 "hakhak": "tertawa", "handak": "bahan peledak", "hansip": "pertahanan sipil",
 "hi": "hai", "hlng": "hilang", "hnya": "hanya", "houm": "rumah", "hrs": "har",
 "hubad": "hubungan angkatan darat", "hubla": "perhubungan laut", "huft": "me",
 "humas": "hubungan masyarakat", "idk": "saya tidak tahu", "ilfeel": "tidak s",
 "imoet": "imut", "info": "informasi", "itung": "hitung", "isengin": "bercand",
 "iyo": "iya", "jablay": "jarang dibelai", "jadul": "jaman dulu", "jancuk": "
 "jdikan": "jadikan", "jg": "juga", "jgn": "jangan", "jijay": "jijik", "jkt": "
 "jth": "jatuh", "jurdil": "jujur adil", "jwb": "jawab", "ka": "kakak", "kaba",
 "kacian": "kasihan", "kadit": "kepala direktorat", "kaga": "tidak", "kaka": "
 "kamtib": "keamanan dan ketertiban", "kamuh": "kamu", "kamyu": "kamu", "kapt",
 "kasat": "kepala satuan", "kasubbid": "kepala subbidang", "kau": "kamu", "kb",
 "kcian": "kasihan", "keburu": "terlanjur", "kedubes": "kedutaan besar", "kek",
 "keliatan": "kelihatan", "keneh": "masih", "kepikiran": "terpikirkan", "kepo",
 "kere": "tidak punya uang", "kesian": "kasihan", "ketauan": "ketahuan", "keu",
 "khan": "kan", "kibus": "kaki busuk", "kk": "kakak", "klian": "kalian", "klo",
 "klwrga": "keluarga", "kmari": "kemari", "kmpus": "kampus", "kn": "kan", "kn",
 "kog": "kok", "kompi": "komputer", "komtiong": "komunis Tiongkok", "konjen": "
 "kpd": "kepada", "kptsan": "keputusan", "krik": "garing", "krn": "karena", "
 "ktny": "katanya", "kudu": "harus", "kuq": "kok", "ky": "seperti", "kykny": "
 "lambreta": "lambat", "lansia": "lanjut usia", "lapas": "lembaga pemasyaraka",
 "lekong": "laki-laki", "lg": "lagi", "lgkp": "lengkap", "lht": "lihat", "lin",
 "lmyan": "lumayan", "lngkp": "lengkap", "loch": "loh", "lol": "tertawa", "lo",
 "lowh": "kamu", "lu": "kamu", "luchu": "lucu", "luff": "cinta", "luph": "cin",
 "maaciw": "terima kasih", "mabes": "markas besar", "macem-macem": "macam-mac",
 "maen": "main", "mahatma": "maju sehat bersama", "mak": "ibu", "makasih": "t",
 "malu2in": "memalukan", "mamz": "makan", "manies": "manis", "mantep": "manta",
 "mba": "mbak", "mending": "lebih baik", "mgkn": "mungkin", "mhn": "mohon", "
 "milis": "mailing list", "mksd": "maksud", "mls": "malas", "mnt": "minta", "
 "mokat": "mati", "mosok": "masa", "msh": "masih", "mskpn": "meskipun", "msng",
 "muahal": "mahal", "muker": "musyawarah kerja", "mumet": "pusing", "muna": "
 "munaslub": "musyawarah nasional luar biasa", "musda": "musyawarah daerah",

```
"nal": "kenal", "nangis": "menangis", "naon": "apa", "napol": "narapidana po
"narsis": "bangga pada diri sendiri", "nax": "anak", "ndak": "tidak", "ndut"
"nekolim": "neokolonialisme", "nelfon": "menelepon", "ngabis2in": "menghabis
"ngambek": "marah", "ngampus": "pergi ke kampus", "ngantri": "mengantri", "n
"ngaruh": "berpengaruh", "ngawur": "berbicara sembarangan", "ngeceng": "kump
"ngeh": "sadar", "ngekos": "tinggal di kos", "ngelamar": "melamar", "ngeliat
"ngemeng": "bicara terus-terusan", "ngerti": "mengerti", "nggak": "tidak", "
"nginep": "menginap", "ngisi": "mengisi", "ngmg": "bicara", "ngocol": "lucu"
"ngumpul": "berkumpul", "ni": "ini", "nyasar": "tersesat", "nyariin": "menca
"nyiram": "menyiram", "nyok": "ayo", "o/": "oleh", "ok": "ok", "prika": "pe
"psn": "pesan", "psti": "pasti", "puanas": "panas", "qmo": "kamu", "qt": "ki
"raskin": "rakyat miskin", "red": "redaksi", "reg": "register", "rejek": "r
"reskrim": "reserse kriminal", "sni": "sini", "somse": "sombong sekali", "so
"sospol": "sosial-politik", "sowry": "maaf", "spd": "sepeda", "sperti": "sepe
"stelah": "setelah", "subbag": "subbagian", "sumbangan": "sumbangkan", "sy":
"tabanas": "tabungan pembangunan nasional", "tar": "nanti", "taun": "tahun",
"te2p": "tetap", "tekor": "rugi", "telkom": "telekomunikasi", "telp": "telep
"tengok": "menjenguk", "terbitin": "terbitkan", "tgl": "tanggal", "thanks":
"thd": "terhadap", "thx": "terima kasih", "tipi": "TV", "tkg": "tukang", "tl
"tman": "teman", "tmbh": "tambah", "tmn2": "teman-teman", "tmph": "tumpah",
"togel": "toto gelap", "tp": "tapi", "tq": "terima kasih", "trgntg": "tergan
"cb": "coba", "y": "ya", "munfik": "munafik", "reklamuk": "reklamasi", "sma"
"ngehe": "kesal", "mz": "mas", "analisis": "analisis", "sadaar": "sadar", "
"nmenarik": "menarik", "zonk": "bodoh", "rights": "benar", "simiskin": "misk
"hardcore": "keras", "akhirx": "akhirnya", "solve": "solusi", "watuk": "batu
"masy": "masyarakat", "still": "masih", "tauk": "tahu", "mbual": "bual", "ti
"ngentotin": "senggama", "kentot": "senggama", "faktakta": "fakta", "sohib":
"trlalu": "terlalu", "nyela": "cela", "heters": "pembenci", "nyembah": "semb
"ikon": "lambang", "light": "terang", "pndukung": "pendukung", "setting": "a
"next": "lanjut", "waspadalah": "waspada", "gantengsaya": "ganteng", "parte"
"nipu": "tipu", "ktipu": "tipu", "jentelmen": "berani", "buangbuang": "buang
"kurng": "kurang", "ista": "nista", "less": "kurang", "koar": "teriak", "par
"problem": "masalah", "tahi": "kotoran", "tirani": "tiran", "tilep": "tilap"
"tak": "tidak", "penertiban": "tertib", "uasai": "kuasa", "mnolak": "tolak",
"taik": "tahi", "wkwkww": "tertawa", "ahokncc": "ahok", "istaa": "nista", "b
"mgkin": "mungkin"}
```

```
processed_comments = []
```

```
for sentence in data['customer_isi_komen']:
    # Remove all the special characters
    processed_comment = re.sub(r'\W', ' ', str(sentence))

    # Converting to Lowercase
    processed_comment = processed_comment.lower()

    #Remove number
    processed_comment = re.sub(r'\d+', ' ', processed_comment)

    # remove all single characters
    processed_comment = re.sub(r'\s+[a-zA-Z]\s+', ' ', processed_comment)

    #remove duplicate character
    pattern=re.compile(r"(\.)\1{1,}",re.DOTALL)
    processed_comment=pattern.sub(r"\1",processed_comment)

    #Corrected Slang words
    words = processed_comment.split()
    rfrm=[slangs[word] if word in slangs else word for word in words]
    processed_comment= " ".join(rfrm)

    #remove stopwords
    factory = StopWordRemoverFactory()
    f = open("stopwords-tala.txt", "r")
```

```

more_stopword = [] #menambahkan stopwords
for line in f:
    stripped_line = line.strip()
    line_list = stripped_line.split()
    more_stopword.append(line_list[0])
f.close()
stopwords = factory.get_stop_words() + more_stopword
temp = [t for t in re.findall(r'\b[a-z]+--[a-z]+\b', processed_comment) if t not
processed_comment = ' '.join(temp)

#stemming
stemmer = StemmerFactory().create_stemmer()
processed_comment = stemmer.stem(processed_comment)

#Subtitusing multiple spaces with single space
processed_comment = re.sub(r'\s+', ' ', processed_comment, flags=re.I)

```

In [5]: processed_comments

Out[5]: ['salah baik fasilitas sedia pilih kamar kamar tempat luas bersih suka jalan pagi ke liling sdh lelah area pandang rooftopnya bagus sarapanya variasi cari lokasi makan te rima kasih layan bagus',
'mobil delica java heritage keluarga endah hernawan hajar hernawan audrey vialita h ernanda aldy hernanda fabian hernanda rudi driver terima kasih rudi pecat sopir fabi an fod and travel asia',
'layak pilih kunjung kota lokasi strategis kamar nyaman kolam renang aman dalam jad i layak pilih',
'juara lokasi keliling makan juara lot parkir luas interior instagramable nuansa ja wa kamar deluxe luas banding kelas pilih makan lengkap lezat staf ramah responsif at as keluh',
'megah bangun fasilitas tunjang kadangkala diri java heritage suguah lengkap megah l engkap fasilitas gedung kerapihan kolam renang view pandang lepas alam balkon lantai restonya pandang indah gunung selamat nan agung nun belah utara sejuk udara khas men u makan tradisional unik menu modern lezat jadi suasana sarap tambah semangat',
'very nice and friendly staf sat sambut security wanita lupa nama sigap sambut seny um bantu bawa koper kesan tahan recomded at saran rawat alat lengkap fasilitas show er kamar tidur sayang rusak extra bed nyaman tipis rangka ranjang sangat karat kotor bawa kamar gudang simpan mesan laku minggu nya siap kamar extra bed nya sehinga tamu betapa kotor rangka extra bed',
'inao java heritage tanggal mei sat check in alam unik mes kamar situs tiket com sam pat ikut undi ambil bola bola kamar luas nyaman heritage interior kamar tebuat bahan kayu unik antik sat buka puasa menu takjil sedia loby makan sahur sedia menu makan t radisional western fasilitas sangat lengkap kolam renang gym alat lengkap sauna jav a heritage pilih inap',
'kamar luas kolam renang bagus luas fasilitas fitnes sauna komplit bersih sarap pil ih enak jatuhcinta jamu dimsum ikan dory nya view gunung slametnya tugas ramah helpf ul security lobi turun bantu mobil tempo time transit dipurwokerto wajib nginap teri makasih java heritage perhati buat kakak perempuan kamar semut over al bagus tips ca ri kamar viewnya gunung kolamrenang iya',
'bagus kamar luas pilih makan pagi mewah kolam renang luas letak strategis hindar r amai ideal libur keluarga sahabat',
'salah sukan sayang inap kamar non smoking pakai tamu smoking ragu biar tamu rokok kamar non smoking breakfast nya sat enak nya kecewa kamar hadap jalan berisik',
'java heritage sekian jamur wilayah parkir luas lobynya homy banget ranah kodam ren ang jaga fresh bersih kamar luas banget enak trus breakfast nya komplit plit menu ta mbah pandang utuh gunung slamet penampakanya',
'inao malam dlm rangka mudik nyaman fasilitas kamar spot instagramable hotel karyaw anyah ramah welcome drink gratis parkir masuk puasa sarap ganti sahur menu sahur luma yan muas gak alun gor sayang gak renang puasa overal is perfect',
'nyaman inap ruang tua lokasi strategis kota parkir luas proses cek in cek out cepa t reseptionist sulit karyawan ramah rom bersih harga relatif mahal kamar ac shower f ungsi kamar kamar mandi bersih wifi cenderung lambat tv chanel chanel anak anak brea

kfast agam standart overal puas sih terimakasih',

'sempat inap gitu kamar mandi sih nga bath tub nya kamar bagus kasur pilih sarapany a banget pengen ambil bawa rumah kolam renang hotel bagus',

'java heritage baik milik fasilitas swiming pol luas bersih kamar relatif fasilitas bintang menu sarapan pagi lengkap employe ramah parkir kendaraan luas',

'java heritage pilih utama inap pergi enggak reservasi kamar akses swiming pol kamar keluarga suka foto sana hasil bagus suka ajak teman fotoin boleh foto foto stay tegu r sopan olah numpang foto sekian tingali kali laku senang foto foto ambil posting so cial media bagus promosi putus inap java heritage',

'didn expect has huge like java heritage the rom is spacious clean the king size be d is really comfortable the breakfast is rich hotel mewah layan ramah kamar luas bers ih wifi gratis sarap pilih jajan pasar omelet hinga pilih real overal had lovely sta y at java heritage',

'java heritage keren kamar luas bersih kolam renang bersih menu sarap enak gak pula ng deh inap nya java heritage gak nyesel deh terimakasih java heritage',

'lokasi strategis dwngan pusat kota masuk area loby sambut tugas ramah kamar ukur s tandar milik luas atas nyaman libur keluarga',

'letak pusat kota milik fasilitas nyaman pusat kuliner jajan hinga butuh inti tawar nyaman pusat kota sahabat',

'lokasi pusat kota pusat kuliner kesan mewah bukti suasana nyaman interior apik lay an tugas ramah bantu cepat fasilitas lengkap kolam renang nyaman makananya enak vari asi rekomendasi',

'salah mewah bintang fasilitas lengkap nyaman setipa ujung pilih inap artis letak menit stasiun fre parking fre wifi fod drink bicycle rental additional charge',

'nama dynasty kali inap ganti nama horison java heritage salah lebih kamar luas lan git langit tingi senang early check in sarap pilih roti kue bubur tradisional jamu t elur real buah roti cane churos nasi limas lauk sederhana enak ruang makan luas lant ai atas lantai pandang gunung slamet tudung awan kolam renang keren senang diskon',

'kali stay java heritage lokal baik turut lokasi strategis dekat enak banget tempat lounge nya rooftop restaurantnya gym swiming pol air pol asin ga enak renang kamar ga usah kamar nya gede bersih jendela lebar ga rom balcony wifi lumayan kenceng breakfa st nya wuenak banget sushi',

'salah lama ronavasi total ganti nama salah baik kolam renang nya ber angin kamar l uas sayang nya kasur nya bunyi pindah badan sarap nya alu lengkap lumayan pilihanya pilih libur',

'inap desember malam proses check in cepat hinga kamar kamar kolam renang kamar lua s megah bersih fasilitas lengkap welcome drink pakai hinga wib menu sarap agam ena k',

'kamar hotel jendela pandang parkir cotage kolam renang standart kamar safe deposit box tv flat chanel batas lemari es isi bayar air mineral botol ukur lengkap mandi se dia lengkap kamar mandi shower makan pagi lantai pemandanganya kolam renang teras du duk nikmat pandang arah gunung makan pagi agam makananya enak wifi kamar lambat',

'kali nginap hotel ambar kondisi pelayananya sempurna bersih layan layan staf ramah menu sarap pagi komplit enak pilih sulit pindah fbsindonesia indonesiantraders ale k',

'java heritage jahe alamat jl dokter angka no sukonegara timur jawa bintang fasilit as ditawarkan kamar bersih bagus modern kamar mandi bersih bagus modern resto bagus d esignya masakanya enak menu nya fariatif saji parkir luas ruang rapat vendor menguna kan ruang rapat ada acara skala pelayananya bagua muas kecewa kunjung bisnis wisat a',

'rekomendasi coba rekomendasi utama kota bagus nyaman bersih staf nya helpfully maka n nya agam menu nya lezat',

'bicara kota lintas dinginya kota mandi air panas lepas inap java heritage konsep b agus kamar balcony pandang gunung lantai dasar lepas teras langsung akses kolam rena ng sofa teras santai sisi makan rooftop suasan makan cozy engan anjak ruang pantas pe nuh kenang memory akses kolam renang kenyam lama situ lokasi parkir luas sehinga fas ilitas',

'kamar bersih luas kulkas kakak perempuan kopi ketel masak air panas botol minum ka mar mandi luas bersih tidur nyaman staf ramah makan variasi enak kolam renang safe d eposit box',

'java heritage salah premium kota kaki gunung slamet gunung rap tingi indonesia wis atawan ujung salah wisata kenal baturaden jarak dukung desain kamar modern fasilita s agam kolam renang pol side bar restaurant rekreasi sempurna malam chekin jam malam jam pagi mesti chek out rasa enak nginap hotel menu sarapanya enak jus',

'lokasi bawah kaki gunung slamet background gunung slamet halaman luas tanam sekita
r segar indah akan tambah nyaman satu',

'kamar nyaman bersih kunjung catat bersih area lorong kamar karpet nya laku pemelih
aran intensif kes awat kes kotor menu breakfast perhati kunjung perhati susun menu b
reakfast nya kualitas layan tahan ramah tamah tamu',

'inap diri ganti nama kali luas banget bangun tingi julang areal luas fasilitas kes
anya mewah elegan kental angkat nilai lokal milik spot cantik foto gak heran jadi ja
va heritage venue foto pre wedding fasilitas lengkap minus gym kemarin sat kabar read
y lebih sih lounge bar cafe restaurant balrom convention center kamar luas ukuranya
kelas fasilitas pada banget',

'tengah kota fasilitas lengkap kolam renang luas ruang temu dl parkir luas sekitar
timur pilih makan nikmat pilih',

'proses cepat sat cek in fo staf ramah kamar nyaman bersih suasana tenang to be con
cern sarap menu variatif perhati kualitas makan lengkap kamar mandi bathfoam minta o
veral layan ramah informatif lini',

'domisili kota kali ganti nama kenal dynasti an ganti horizon ganti nama java herit
age salah mewah kota mendoan sejajar aston santika dl renang kolam renang air hangat
soft drink gratis',

'java haritage mewah manja ujung bangun nuansa haritage salah uni bangun eks bangu
n horison',

'hotel nyaman kelas bintang bagus pelayanannya ramah sopan kamar bersih kolam renang
god cocok olahraga libur makan cofe breaknya variatif overal god',

'inap java heritage layan dapat muas staycation pegawai sigap ramah welcome drink p
roses check in cepat tempat kamar deluxe kamar luas banget nyaman inap debit air der
as kamar kamar mandi bersih ranjang luas serba luas swimming pool bikin betah lama ren
ang cafe swimming pool sarap muas variasi makan limpah lezat kurang terang sekitar lor
ong redup rekomendasi inap hotel',

'suka layan bagus layan restoran layan amplas jogja makan kamar bagus direcomend te
man teman check in kemarin sabtu check out',

'pilih java heritage option inap aston star hubung perfect relax pilih stay frontde
sk pesan hari pas dtv via agoda rate mahal pesan langsung pesan kamar delux sat chec
kin double bed bilang gada full checkout sore enga mrk bilang enga pas malam pergi pe
san taksi tamu tawar kamar delux twin bed double bed such weird room overal oke cing
tingi kamar luas star ga cm ksh handuk kasih handuk',

'kali inap htl nama htl group internasional nasional mutu standard htl kota sat pel
ayanannya ramah sahabat lobynya hangat asa kota kamar interior kamar mandi bersih mod
ern menu ruang sarapnya enak santai lama area buka kursi outdoor taman kolam mampir
rencana malam',

'inap senang staf sambut datang tamu ramah proses check in check out cepat interior
modern mewah kamar luas fasilitas bagus menu sarap aneka ragam enak kesan lidah bubu
r sumsum aneka pastry',

'inap salah solusi cari asik segi tata ruang adem menu sarap pilih kualitas kamar n
yaman lengkap kursi santai nonton tv nyaman pilih duduk kasur ac dingin saran kasih
timer pas pagi dingin banget maklum orang kampung nyobain kolam renang hari inap pil
ih hiburan sedia belah tinggal jalan kaki nyampe parkir luas jamin aman pas malam sekur
iti keliling overal oke keluarga jalan kerja minus kamar pingir hadap jalan',

'malam java heritage senang alam nikmat gaya indonesia modern lupa sempat sarap jam
pagi indah pandang kota alam indah santap',

'inap saudara meni mewah bintang kota hubung malas kendra mobil kereta hotel sedia
jemput stasiun stasiunya tempat telepon gengam operator lte masuk lobi luas kamar lu
as kolam renang buesar banget kolam anak anak pisah balon main balon perosot lorong
lorong samping kolam renang check in kecewa pesan muka catat bayar muka masuk kamar
cek keliling kamar luas harga amenities nya pelit jujur deh shampo sabun batang loti
on',

'loby pas desember dekorasi natal bagus samping loby swimming pool duduk tuk santa ka
mar luas bersih nyaman shower kamar mandi tuk anak',

'hilang handphone loby akses ctv kamera aju komplain staf konsisten sore telpon ctv
telpon memperlihatkan ctv curiga ctv rekam sehinga tunjuk tamu ketidakkonsistenan pr
ofesional jaga aman tamu inap aman nyaman tamu prioritas nomor sekian inap malam tin
gkat ocupansi tingi sewa lapang parkir event',

'nama horizon ganti milik java heritage ruang kamar harga lumayan jangkau stasiun l
okasi makan',

'letak strategis kamar luas nyaman view kolam renang gunung fasilitas kamar kamar m
andi bagus suka sandal hotel ahahaha kolam renang luas harga menu cafe mahal menu br

eakfastnya juara enak pilih tempe mendoan khas enak proses check in check out cepat terima kasih java heritage',

'ganti nama java heritage kesan comfort banget parkir luas kamar lumayan bersih sayangnya pas inap sana keran air panas fungsi stafnya ramah staf kolam renang sebal sih pusat kota kolam renang nya bersih pas kesini weekday ga ramai deh kolam nya renang bebas rekomendasi cari nyaman',

'nuansa resort kota nama java heritage kamar luas milik kolam renang luas sarapannya enak variasi saji kuliner khas tempe mendoan soto soto kuah bumbu kacang alam senang cek in cek out',

'nuansa resort kota kaki gunung slamet jawa kanarnya luas saji makan pagi konversi makan sahur ramadhan menu menu asin saji makan an variatif',

'pilih horizon timbang kolam renang anak kolam nya luas anak enjoy ruang cium aroma segar makan nya pilih enak tugas nya ramah tawar bantu kamar bersih luas pokoke beta deh',

'lokasi strategis kota bersih tata rapi renovasi kolam renang kamar bungalow pisah bangun sayang sedia menu sarap vegetarian populer',

'pilih horizon letak kota mudah pergi staf ramah harga kamar relatif terjangkau breakfastnya enak variasi balkon kamar langsung view pool senang',

'makan enak kamar bersih luas layanan ramah strategis halaman parkir luas cocok acara ada acara undang serta harga lumayan standar',

'kamar balkon menghadap kolam renang bonus pandang gunung slamet selimut awan kilas baturaden lereng kondisi matahari pagi cerah malam hujan',

'satu hadir seminar horizon suasana nyaman pusat kota parkir luas kolam renang bersih bagus kamar luas nyaman bersih oia wifi nya lancar nice one',

'asal kota makasar rangka hadir lustrum fakultas terna universitas jenderal soedirman check in tanggal feb kamar luas nyaman ukur deluxe lengkap fasilitas kolam renang luas serta cafe resto pingirnya sehingga nyaman lama renang santai kolam tata rapi mewah elegan lighting tawar malam cocok minum kopi okok menu sarap lezat serta makan khas mendoan jamu tradisional jawa kunyit asam temu lawak fasilitas atas tamu seru asyik',

'wisata batu rade lereng gunung slamet belah selatan istimewa mandi air panas jalan lelah madi bayar selup kaki mandi air hangat satu area kebun binatang rekreasi edukasi anak',

'libur purwokerto inap horizon purwokerto layanan ramah staf bantu kamar sedia mini bar kamar tempat swimming pool garden parkir kendaraan luas penting lokasi strategis pusat kota',

'warga salah inap horizon ultima nama dynasty ubah horizon tempat bagus kolam renang bentuk beda lobby enak senang',

'inap horizon tanggal desember anak bekas dynasty hotel tata kecewa lantai kamar karpet sikat gigi orang sampo kesanya pelit restaurant bangun puas lantai kamar karpet give',

'tahap renovasi keluarga inap sehingga kolam renang satu sarap pilih menu tugas ramah anak ajak warna hias donat satu selesai sarap tugas resepsionis satu check in talu ramah kamar luas nyaman interior elegan tidur empuk kamar connecting door tutup bidang kasur pintu sehingga suara keluarga kamar belah dengar kamar koridor lift kamar sepi sehingga timbul kesan spooky kosong tahap renovasi nyaman',

'milik ekspektasi oini layanan memuaskan inap kamar deluxe with balcony kamar luas ceiling tinggi kesan luas shet sarung bantal bersih wangi kasur bantal empuk pelit amenities kamar mandi luas koneksi wifi kecewa habis kamar hingga lobby akses putus makan sarap makan siang malam menggunakan ruang meeting enak variasi sarap temu makan japanese indonesia chinese western inap iya layanan jemput stasiun gratis drop stasiun',

'hanimun dadak sengaja pilih cek in tanggal sibuk kamar cancel an layanan resepsionis sayang belboynya sigap jalan cotage siksa high heels bawa barang tukang parkir sehingga supir sulit parkir di dekat lobby checkout jalan lobby jarak lumayan kecewa sih momen baik nya kaki lecet high heels hiks',

'kamar bed nyaman sedia smoking floor harus breakfast sesuai harap mendoan khas sayang lokasi pusat kota proses booking kendala alamat email habis kuota sehingga email konfirmasi travel agent kendala kirim ulang email bouncing muncul mailer daemon bla bla sehingga lambat cek in staf bantu',

'bersih nyaman ramah point semat senyum khas banyumas driver jemput langsung sambut stasiun jarak menit jalan menggunakan mobil kamar bersih elegan makanannya lumayan enak suka sepeda sedia moda transportasi kayuh roda gratis lho',

'jelang malam staf layanan sigap ramah satu tunggu check in staf antar welcome drink hangat campur kakak perempuan madu lemon nikmat proses check in cepat langsung antar k

amar nyaman jendela hadap jalan oya suka selasar lebar temaram interior kamar mewah biarpun detail oke sayang salah handuk bau apek dekil lawan handuk wangi putih staf housekeeping teliti inap mengangu lai oke sarap',

'jalan lumayan dipurwokerto lumayan mudah menemukanya nama milik cabang dynasty amb il alih jaring horison bangun utama bangun modern kemarin inap tipe cotage deret kamar milik teras sendiritepat bangun utama bangun bangun peninggalan dynasty kamar dik sepeda ontel parkir rapi depan loby kitabebas mengunakanya kendaraan pribadi langsung parkir depan kamar kondisi kamar kondisi tipe shet trlple shet bathtub',

'sdg tahap renovasi an foto sdh kolam renang selesai bangun september informasi res rpsionis kamar menghadap gunung lbh balkon hadap jalan raya balkon kecoh harga tawar agoda search engine lai kenyat nya reservasi langsung by phone lbh murah informasi kamar lbh akurat htl jual type kamar jenis cotage notabene sdh jaman sbml pesan agoda apa baik cros check reservasi nya',

'kamar bersih luas lengkap karyawan sapa senyum sat inap sdg manager pusat ramah menu makan lengkap variasi saran koran lokal lbh opsional koran nasional',

'inap horison sabtu juni kondisi bangun utama nama dynasty tahu rom cotages deluxe lokasi samping bangun utama lengkap menggunakan nama dynasty cangkir kamar merica garam malabar restaurant menggunakan nama dynasty kondisi kamar luas kasur ukur king size lengkap bath tube bangun interior sarap pagi lokasi samping gedung utama malabar restaurant menu',

'pesan kamar via boking com cek in mlm pesan kamar harga rendah kamar cotage parkir mobil langsung kamar enak sih kecewa spreinya noda lusuh dinding kamar noda redam pintu conecting buruk salur tvi ga bagus gambar harga murah hasil murah ga jamin nama brand nama pikir harga promosi ga yah pelang kecewa terimakasih maaf',

'pelayananya bersih inisiatif bersih datang rom service lai iya',

'ckup lokasi strategis layan muas service customer carenya bgtnu cepat bilang bangun cermin globalisasi cermin letak lingkung jawa lbih adat timur',

'bagus wifi kencang kamar bersih suruh sdh pencet make up rom breakfast makan enak maners letak bilang bagus swalayan moro mohon baik bersih kamar suka lambat',

'pugar nama dynasty bangunanya pugar megah mas menggunakan perabot sofa nya kalu kamar uda bagus bersih oke karpet jalan nya mas pakai karpet kualitas inap bilang kualitas bintang',

'bintang tahap renovasi aston baik sat horison kejar ketertinggalanya aston renovasi kamar nyaman sat check in pilih kamar renovasi variasi menu breakfast masakanya enak',

'kamar mandi kelas melati awat ac jelek dingin sarung bantal seprai usang alias buluk kutu busuk kamar selimut tipis banget kelas melati kamar hotel kecewa sesuai harga tinggi tawar',

'inap malam keluarga kes surprise modern kota maklum renovasi kamar nya bagus nyaman design rom okay star sayang wifi restaurant benah inap',

'arik coba tawar promo salah situs konsep kamar bangun kamar letak luar sekitar hal aman langsung parkir belah kamar garasi mobil bawa kendaraan paksa jalan kaki iya kemarin renovasi jala muter kamar lumayan bersih luas aga kecewa kamar mandi luas tua luar bangun suka makananya murah enak pagi coba breakfafst bufet nya option nya',

'tanggal feb kmaren inap sana layan ramah ah asa tamu vip gitu nyata tsb layan jempu t stasiun biaya rb dng menggunakan avanza kolam renang sepuh rampung keramik lantai dasar kolam beranta aman sih sat renang breakfastnya enak lumayan komplit jamu sala d',

'tepat acara porprov panjang pekan olahraga provinsi sehinga penuh kamar horison kamar kamar bilang luas sofa meja bathtub kamar gedung kesanya jaman gedung sinyal kamar jelek hp wifi kamar sinyal hp sinyal wifi solusi suka nongkrong loby untung loby hotel luas nyaman tekan masakan enak sarap menu sarapanya variatif',

'tua pondok baik aston bangun gambar lihat gambar profil rencana kembang proses pem bangunanya selesai bagus bagus bangun dengar suara berisik malam sat skor sarapanya lumayan air panas masalah hangat panas kolam renang tidur king koil bawa mobil parkir kamar',

'puas harga sesuai fasilitas terima sarap gratis hidangan variasi baik jalan ujung asia tengara kes nyanyi restoran sat malam salah nyanyi perempuannya nyanyi blues bagus aksi lounge jenis las vegas inap',

'pilih bintang inap malam juluk baik kota kecewa harap tua dynasty kamar bersih furniture tua butuh renovasi sulit capai lokasi untung gps handal tunjuk jalan jalan bl okir pemeliharaan sehinga ambil rute alternatif sarap kesan standar restoran bintang lalat liar bebas meja kotor habis kadang kadang renang kolam renang ubin kolam renang rusak ekstra hati hati aspek positif',

'transit jakarta bersih tarik harga murah banding bintang bandung merekomendasikanya',

'lobi modern mewah kamar luas tidur nyaman sarap pada satu minus handuk kuning uba h',

'bagus lokasi bagus pusat kota daerah baturaden daerah wisata utama kamar luas bersih kamar mandi bersih sarap',

'inap malam pilih bungalow kolam renang kamar tua bersih staf bantu makan baik kolam renang bersih nikmat inap',

'saudara kejut kamar madu waris budaya jawa bagus interior layan kecewa sat kue madu salam surat nama salah surat fatal hubung reservasi tindak maf',

'kali inap lobi area kolam renang bagus indonesia ramah staf orang sambut pintu orang pekerjaan makan restoran sisi mahal burger kentang goreng kentang goreng ngomong-ngomong makan prasmanan aneh ragu rekomendasi orang perspektif bisnis wifi bagus kolam renang undang',

'terima kasih staf waris jawa layan sempurna waris budaya jawa milik kamar pandang bagus restoran bagus kolam renang bagus waris budaya jawa rekomendasi orang orang unjung',

'waris budaya jawa bagus suka suasana musik gamelan jawa staf senang bantu kolam renang staf temu kamar tidur kamar mandi bersih gantung pakai kamar mandi suka makan mes sup buntut sup wonton mie goreng lezat sarap lengkap saji telur rekomendasi inap unjung',

'rekomendasi pilih daerah staf senang manajer bagus helpful pandang bar gunung slamet langit unjung',

'sedia aula ada dekat layak habis uang sewa ruang aula jalan temu kolam renang',

'tanda tanda manajemen baik detail lukis dl tanda tanda pergi pakai air mata aneh bicara staf sarap jalan timur pilih makan layak kolam renang bagus buruk kecewa waris lai istimewa pikir harga waris bayar harga mewah kota',

'pergi desember mes kamar libur keluarga lokasi kenal horison kesan fa ade modern indah staf ramah parkir kamar luas milik langit langit tingi dianggap area bagus sempurna jaga fasilitas awat meja kamar gores anak anak cinta kolam renang kuliner menit kendaraan ruang baik menu sarap pagi makan dapat rempah rempah kamar bagus lokasi sempurna bintang',

'bagus harga kamar kamar mandi kolam renang dl gaya klasik pilih menu sarap staf bantu salah baik kota tawar pandang indah gunung rekomendasi',

'unjung waris jawa natal kamar luas parkir makan minum variasi sarap manajemen lantai karpet koridor kotor',

'letak kota saing kota besar provinsi jawa jam jalan kereta eksekutif jakarta horison manajemen jawa waris budaya kota milik bandara bandara dekat bandara jogjakarta cilacap proses check in check out cepat staf bantu bantu atur lobi ruang pisah sofa nyaman okok inap kamar premiere balkon hadap kolam renang gunung slamet luas bersih awat ac dingin santa iklim panas tidur nyaman fasilitas lengkap tv',

'kenal horison letak pusat kota simpang batu raden aston milik parkir sehingga khawatir bawa mobil milik kamar luas pilih kamar sambung nd lantai pandang kota parkir milik kamar pandang kolam renang kolam renang anak anak orang dewasa restoran belah kolam renang damping sarap restoran letak lantai variasi makan lengkap pagi nasi goreng bubur ayam soto churos pancake sushi pilih ayam goreng',

'lihat foto ulasan ringkas pilih malam kamar luas tipe deluxe karpet bersih nyaman main piano lobi suasana inap nyaman kecewa singkat set gamelan main jadwal pintu sambutan ramah pikir staf meja informasi turis pikir kota tarik wisatawan',

'salah waris keliling pohon romantis pandang kota lantai manja staf ramah dengarnya anyi burung pagi temu isi majalah lokal prasmanan coba dim sum pulih tenaga renang kolam renang keras minggu pagi',

'inap malam ramah kamar luas parkir kolam renang bagus restoran layak saji hidangan standar sarap kenal soto lihat foto staf layan',

'lebih luas bersih kamar nyaman bangun koneksi internet wifi tulus cari receptionis akses mudah stasiun kereta api pusat kota hidup malam check in cepat check out menu sarap standar parkir luas henti hiburan karaoke klub malam cafe kurang mahal rekomendasi inap horison bangun favorit kota saran tinggal cotage bangun tua ac perabot tua kamar mandi kotor',

'inap malam kamar tipe deluxe cotage kamar bangun penuh kesan kamar bersih fungsional kaca kotor cangkik menggunakan dynasty handuk keset layak keran wastafel kotor untuk sarap tingkat suasana hati makanannya variasi bagus campur sari kelompok restoran saran anti properti cotage menggunakan properti formerly dinasti',

'inap renovasi mengganggu kembang favorit kali unjung kamar favorit bungalow kolam re

nang staf ramah hiburan malam hiburan harap tampil renovasi selesai',
 'ujung rutinitas anak ajar ujung pekan inap horizon rd pilih kali coba inap kamar gedung milik pandang lupa kolam renang view kamar kolam renang bar mudah temu parkir sarap temu sarap gaya amerika gaya jawa cina makan jepang inap',
 'kali tinggal horizon dinasti kali tinggal tampil konsep kolam renang bentuk bujur sangkar kamar pikir wifi internet gratis cepat sayang restoran saran anti sabun sabun cair inefficient sabun tamu menggunakan',
 'pilih bangun tua indah masuk ruang nyaman ac tua lampu meja bising furnitur tua putus mengupgrade kamar kamar deluxe premium balkon kamar luas lantai parket butuh awat tidur bantal bagus kamar bersih sedia fasilitas ering rambut selesai staf ramah informasi arah kamar dl sarap bangun kamar agam makan lezat gratis lambat check out kerajaan harga dapat resepsionis',
 'kamar luas nyaman layanan cepat check in out renovasi oktober kamar sisi istirahat ubah penguridan kamar langsung sisi lampu tidur waspada musik keras pekan lewat pm pemeriksaan renovasi hiburan pekan bising mengganggu malam hiburan menganggu kamar lengkap besi kirim catat papan setrika',
 'salah baik parkir luas kamar milik kolam renang chers cafe and convention center discotheque pesta nikah sarap',
 'tinggal malam kesan gedung awat staf tahu sedia bantu rekomendasi ujung ruang kamar belah lift milik apa tawar jendela hadap koridor kamar handuk mandi tempat kamar mandi kertas toilet sedia milik apa handuk tangan wajah handuk mandi pinjam ering rambut malam habis stock pemeriksaan']

Vektorisasi

In [6]:

```
from sklearn.feature_extraction.text import CountVectorizer
tf_vectorizer = CountVectorizer(max_df=1.0, min_df=1)
tf = tf_vectorizer.fit_transform(processed_comments)
#hasil representasi
tf_terms = tf_vectorizer.get_feature_names()
print(tf_vectorizer.get_feature_names())
matrix = tf.toarray()
print(matrix)
```

```
['ac', 'acara', 'ada', 'adat', 'ade', 'adem', 'additional', 'aga', 'agam', 'agent',
'agoda', 'agung', 'ah', 'ahahaha', 'air', 'ajak', 'ajar', 'aju', 'akan', 'akses', 'aksi', 'akurat', 'al', 'alam', 'alamat', 'alat', 'aldy', 'alek', 'alias', 'alih', 'alternatif', 'alu', 'alun', 'aman', 'ambar', 'ambil', 'amenities', 'amerika', 'amplas', 'an', 'anak', 'and', 'aneh', 'aneka', 'angin', 'angka', 'angkat', 'anjak', 'antar', 'anti', 'antik', 'apa', 'apek', 'api', 'apik', 'arah', 'area', 'areal', 'arik', 'aroma', 'artis', 'asa', 'asal', 'asam', 'asia', 'asik', 'asin', 'aspek', 'aston', 'at', 'atas', 'atur', 'audrey', 'aula', 'avanza', 'awan', 'awat', 'ayam', 'background', 'badan', 'bagua', 'bagus', 'bahan', 'baik', 'balcony', 'balkon', 'balon', 'balroom', 'bandara', 'banding', 'bandung', 'banget', 'bangun', 'bangunanya', 'bantal', 'bantu', 'banyumas', 'bar', 'barang', 'batang', 'batas', 'bath', 'bathfoam', 'bathtub', 'batu', 'baturaden', 'bau', 'bawa', 'bawah', 'bayar', 'be', 'bebas', 'bed', 'beda', 'bekas', 'belah', 'belboynya', 'benah', 'bentuk', 'ber', 'beranta', 'berisik', 'bersih', 'besar', 'besi', 'betah', 'betapa', 'bgtu', 'biar', 'biarpun', 'biaya', 'bicara', 'bicycle', 'bidang', 'bikin', 'bilang', 'binatang', 'bintang', 'bising', 'bisnis', 'bla', 'blokir', 'blues', 'boking', 'bola', 'boleh', 'bonus', 'botol', 'bouncing', 'box', 'brand', 'breakfast', 'breakfast', 'breakfastnya', 'breaknya', 'buah', 'buat', 'bubur', 'budaya', 'buesar', 'bufet', 'bujur', 'buka', 'bukti', 'buluk', 'bumbu', 'bungalow', 'buntut', 'bunyi', 'burger', 'buruk', 'burung', 'busuk', 'butuh', 'by', 'cabang', 'cafe', 'cair', 'campur', 'cancel', 'cane', 'cangkir', 'cantik', 'capai', 'carenya', 'cari', 'catat', 'ceiling', 'cek', 'cekout', 'cenderung', 'center', 'cepat', 'cerah', 'cermin', 'chanel', 'charge', 'check', 'checkin', 'checkout', 'check', 'checkin', 'chers', 'chinese', 'churos', 'cilacap', 'cina', 'cing', 'cinta', 'cium', 'ckup', 'clean', 'cm', 'coba', 'cocok', 'cofe', 'com', 'comfort', 'comfortable', 'concern', 'conecting', 'convention', 'cotage', 'cotages', 'cozy', 'cros', 'ctv', 'curiga', 'customer', 'dadak', 'daemon', 'daerah', 'dalam', 'damping', 'dapat', 'dasar', 'datang', 'debit', 'deh', 'dekat', 'deket', 'dekil', 'dekorasi', 'delica', 'delux', 'deluxe', 'dengar', 'depan', 'deposit', 'deras', 'deret', 'desain', 'desember',
```

'design', 'designya', 'detail', 'dewasa', 'dianggap', 'did', 'didn', 'dik', 'dim', 'd
 imsum', 'dinasti', 'dynasty', 'dinding', 'dingin', 'dinginya', 'dipurwokerto', 'dire
 comend', 'diri', 'discotheque', 'diskon', 'ditwarkan', 'dl', 'dlm', 'dng', 'dokter',
 'domisili', 'donat', 'dor', 'dory', 'double', 'drink', 'driver', 'drop', 'dtv', 'dud
 uk', 'dukung', 'dwngan', 'dynasti', 'dynasty', 'early', 'edukasi', 'ekat', 'eks', 'e
 ksekutif', 'ekspektasi', 'ekstra', 'elegan', 'email', 'employe', 'empuk', 'enak', 'e
 ndah', 'enga', 'engak', 'engan', 'engine', 'enjoy', 'ering', 'es', 'event', 'expec
 t', 'extra', 'fa', 'fabian', 'fakultas', 'fariatif', 'fasilitas', 'fatal', 'favori
 t', 'fbsindonesia', 'feb', 'fitnes', 'flat', 'flor', 'fo', 'fod', 'formly', 'foto',
 'fotoin', 'fre', 'fresh', 'friendly', 'frontdesk', 'ful', 'fungsi', 'fungsional', 'f
 urnitur', 'furniture', 'ga', 'gada', 'gak', 'gambar', 'gamelan', 'ganti', 'gantung',
 'garam', 'garasi', 'garden', 'gausah', 'gaya', 'gede', 'gedung', 'gengam', 'gigi',
 'gitu', 'give', 'globalisasi', 'god', 'gor', 'goreng', 'gores', 'gps', 'gratis', 'gr
 oup', 'gudang', 'gunung', 'gym', 'habis', 'had', 'hadap', 'hadir', 'hajar', 'halama
 n', 'handal', 'handphone', 'handuk', 'hangat', 'hanimun', 'harap', 'harga', 'hari',
 'haritage', 'harus', 'has', 'hasil', 'hati', 'helpful', 'helpful', 'helpfully', 'hent
 i', 'heran', 'heritage', 'hernanda', 'hernawan', 'hias', 'hibur', 'hidang', 'hidup',
 'highels', 'hiks', 'hilang', 'hindar', 'hinga', 'homy', 'horison', 'horizon', 'hote
 l', 'housekeping', 'hp', 'htl', 'hubung', 'huge', 'hujan', 'ideal', 'ikan', 'iklim',
 'ikut', 'in', 'inap', 'indah', 'indonesia', 'indonesiantraders', 'informasi', 'infor
 matif', 'inisiatif', 'instagramable', 'intensif', 'interior', 'internasional', 'inte
 rnet', 'inti', 'is', 'isi', 'istimewa', 'istirahat', 'iya', 'jadi', 'jadwal', 'jag
 a', 'jahe', 'jajan', 'jakarta', 'jala', 'jalan', 'jam', 'jaman', 'jamin', 'jamu', 'j
 amur', 'jangkau', 'japanese', 'jarak', 'jaring', 'jatuhcinta', 'java', 'jawa', 'jela
 ng', 'jelek', 'jemput', 'jendela', 'jenderal', 'jenis', 'jepang', 'jl', 'jogja', 'jo
 gjakarta', 'jual', 'juara', 'jujur', 'julang', 'juluk', 'juni', 'jus', 'kabar', 'kac
 a', 'kacang', 'kadang', 'kadangkala', 'kakak', 'kaki', 'kali', 'kalu', 'kamar', 'kam
 era', 'kampung', 'kanarnya', 'karaoke', 'karat', 'karpet', 'karyawan', 'karyawanya',
 'kasih', 'kasur', 'kayu', 'kayuh', 'kebun', 'kece', 'kecewa', 'kecoh', 'kejar', 'kej
 ut', 'kelas', 'keliling', 'kelompok', 'keluarga', 'keluh', 'kemarin', 'kembang', 'ke
 nal', 'kenang', 'kencang', 'kenceng', 'kendala', 'kendara', 'kendaraan', 'kental', 'k
 entang', 'kenyam', 'kenyat', 'keramik', 'keran', 'kerapihan', 'keras', 'keren', 'ker
 eta', 'kerja', 'kertas', 'kes', 'kesan', 'kesanya', 'keset', 'kesini', 'ketel', 'ket
 ertinggalanya', 'ketidakonsistenan', 'khas', 'khawatir', 'kilas', 'king', 'kirim', 'k
 itabebas', 'klasik', 'klub', 'kmaren', 'kodam', 'koil', 'kolam', 'kolamrenang', 'kom
 plain', 'kompli', 'kondisi', 'koneksi', 'konfirmasi', 'konsep', 'konsisten', 'konve
 rsi', 'koper', 'kopi', 'koran', 'koridor', 'kosong', 'kota', 'kotor', 'ksh', 'kuah',
 'kualitas', 'kue', 'kuliner', 'kulkas', 'kuning', 'kunjung', 'kunyit', 'kuota', 'kur
 ang', 'kursi', 'kutu', 'lai', 'laku', 'lalat', 'lama', 'lambat', 'lampu', 'lancar',
 'langit', 'langsung', 'lantai', 'lapang', 'las', 'lauk', 'lawak', 'lawan', 'layak',
 'layan', 'lbh', 'lbih', 'lebar', 'lebih', 'lecet', 'lelah', 'lemari', 'lemon', 'leng
 kap', 'lepas', 'lereng', 'letak', 'lewat', 'lezat', 'lho', 'liar', 'libur', 'lidah',
 'lift', 'lighting', 'lihat', 'like', 'limas', 'limpah', 'lingkung', 'lini', 'linta
 s', 'lobi', 'loby', 'lobynya', 'lokal', 'lokasi', 'lorong', 'lot', 'lotion', 'loung
 e', 'lovely', 'lte', 'luar', 'luas', 'lukis', 'lumayan', 'lupa', 'lustrum', 'lusuh',
 'madi', 'madu', 'maf', 'mahal', 'mailer', 'main', 'majalah', 'makan', 'makananya',
 'makasar', 'make', 'maklum', 'malabar', 'malam', 'malas', 'mampir', 'manager', 'mana
 jemen', 'manajer', 'mandi', 'maners', 'manja', 'mas', 'masak', 'masakan', 'masakany
 a', 'masalah', 'masuk', 'mata', 'matahari', 'media', 'megah', 'mei', 'meja', 'melat
 i', 'memory', 'memperlihatkan', 'mendoan', 'menemukanya', 'menghadap', 'mengangu',
 'menganguku', 'mengunakan', 'mengunakanya', 'mengupgrade', 'menhadap', 'meni', 'meni
 t', 'menu', 'merekomendasikanya', 'merica', 'mes', 'mesan', 'mesti', 'meting', 'mewa
 h', 'mh', 'mie', 'milik', 'mineral', 'mingu', 'mini', 'minta', 'minum', 'minus', 'ml
 m', 'mobil', 'moda', 'modern', 'mohon', 'momen', 'moro', 'mrk', 'muas', 'mudah', 'mu
 dik', 'muka', 'muncul', 'murah', 'musik', 'muter', 'mutu', 'nama', 'nan', 'nasi', 'n
 asional', 'natal', 'nd', 'nga', 'nginap', 'ngomong', 'nice', 'nikah', 'nikmat', 'nil
 ai', 'no', 'noda', 'nomor', 'non', 'nongkrong', 'nonton', 'notabene', 'nuansa', 'num
 pang', 'nun', 'nya', 'nyaman', 'nyampe', 'nyanyi', 'nyata', 'nyesel', 'nyobain', 'oc
 upansi', 'oini', 'oiya', 'okay', 'oke', 'okok', 'oktober', 'olah', 'olahraga', 'omel
 et', 'one', 'ontel', 'operator', 'opsional', 'option', 'orang', 'out', 'outdoor', 'ov
 er', 'overal', 'oya', 'pada', 'pagi', 'pakai', 'paksa', 'panas', 'pancake', 'pandan
 g', 'panjang', 'pantas', 'papan', 'parket', 'parking', 'parkir', 'pas', 'pasar', 'pa
 stry', 'pasuk', 'pecat', 'pegawai', 'pekan', 'pekerjaan', 'pelang', 'pelayananya', 'p
 elit', 'pemandanganya', 'pembangunanya', 'pemeliharaan', 'penampakanya', 'pencet', 'p

engin', 'pengurdian', 'peninggalan', 'penting', 'penuh', 'perabot', 'perempuan', 'per
empuannya', 'perfect', 'pergi', 'perhati', 'periksa', 'perosot', 'perspektif', 'pesa
n', 'pesta', 'phone', 'piano', 'pikir', 'pilih', 'pilihanya', 'pindah', 'pingir', 'p
ingirnya', 'pinjam', 'pintu', 'pisah', 'plit', 'pm', 'pohon', 'point', 'pokoke', 'po
l', 'pondok', 'populer', 'porprov', 'positif', 'posting', 'prasmanan', 'pre', 'premi
ere', 'premium', 'pribadi', 'prioritas', 'profesional', 'profil', 'promo', 'promos
i', 'properti', 'proses', 'provinsi', 'puas', 'puasa', 'pugar', 'pulang', 'puluh',
'purwekerto', 'pusat', 'putih', 'putus', 'rade', 'raden', 'ragam', 'ragu', 'ramadha
n', 'ramah', 'ramai', 'rambut', 'rampung', 'ranah', 'rangka', 'ranjang', 'rap', 'rap
at', 'rapi', 'rapih', 'rasa', 'rate', 'rawat', 'raya', 'rb', 'rd', 'ready', 'real',
'realy', 'receptionis', 'recomend', 'recomnded', 'redam', 'redup', 'rekam', 'rekomen
dasi', 'rekreasi', 'relatif', 'relax', 'rempah', 'renang', 'rencana', 'rendah', 'ren
ovasi', 'rental', 'resepsionis', 'reseptionist', 'reservasi', 'resort', 'responsif',
'resrpsionis', 'restaurant', 'restaurantnya', 'resto', 'restonya', 'restoran', 'ric
h', 'ringkas', 'roda', 'roftop', 'roftopnya', 'rokok', 'rom', 'romantis', 'ronavas
i', 'roti', 'ruang', 'rudi', 'rumah', 'rusak', 'rute', 'rutinitas', 'sabtu', 'sabu
n', 'safe', 'sahabat', 'sahur', 'saing', 'saji', 'salad', 'salah', 'salam', 'salur',
'sambung', 'sambut', 'sambat', 'samping', 'sampo', 'sana', 'sangan', 'sangat', 'san
gkar', 'santa', 'santai', 'santap', 'santika', 'sapa', 'saran', 'sarap', 'sarapanpag
i', 'sarapanya', 'sari', 'sarung', 'sat', 'satu', 'saudara', 'sauna', 'sayang', 'sbl
m', 'sdg', 'sdh', 'search', 'sebal', 'security', 'sederhana', 'sedia', 'segar', 'seg
i', 'sehinga', 'sejajar', 'sejuk', 'sekian', 'sekitar', 'sekuriti', 'selamet', 'sela
sar', 'selatan', 'selesai', 'selimut', 'selup', 'semangat', 'semat', 'seminar', 'sem
pat', 'sempurna', 'semut', 'senang', 'sendal', 'sendiritepat', 'sengaja', 'senyum',
'sepeda', 'sepenuh', 'sepi', 'seprai', 'september', 'serasa', 'serba', 'serta', 'ser
vice', 'sesuai', 'set', 'setipa', 'setrika', 'sewa', 'shampo', 'shet', 'shower', 'si
ang', 'siap', 'sibuk', 'side', 'sigap', 'sih', 'sikat', 'siksa', 'simpan', 'simpan
g', 'singkat', 'sinyal', 'sisi', 'situ', 'situs', 'size', 'skala', 'skor', 'slamet',
'slametnya', 'smoking', 'social', 'soedirman', 'sofa', 'soft', 'solusi', 'sopan', 's
opir', 'sore', 'soto', 'spacious', 'spoky', 'spot', 'spreinya', 'sroto', 'staf', 'st
andar', 'standard', 'standart', 'star', 'stasiun', 'stasiunnya', 'stay', 'staycatio
n', 'stock', 'strategis', 'suara', 'suasa', 'suasan', 'suasana', 'such', 'suguh', 's
uka', 'sukan', 'sukonegara', 'sulit', 'sum', 'sumsum', 'sup', 'supir', 'surat', 'sur
prise', 'suruh', 'sushi', 'susun', 'swalayan', 'swimming', 'tahan', 'tahap', 'tahu',
'takjil', 'taksi', 'talun', 'tamah', 'taman', 'tambah', 'tampil', 'tamu', 'tanam', 't
anda', 'tangkal', 'tangan', 'tarik', 'tata', 'tawan', 'tawar', 'tebuat', 'tegur', 'te
kan', 'telepon', 'teliti', 'telpon', 'telur', 'teman', 'temaram', 'tempat', 'tempe',
'tempo', 'temu', 'tenaga', 'tenang', 'tengah', 'tengara', 'tepat', 'terang', 'tera
s', 'terima', 'terimakasih', 'terna', 'the', 'tidur', 'tiket', 'timbang', 'timbul',
'time', 'timer', 'timur', 'tindak', 'tingal', 'tingali', 'tingi', 'tingkat', 'tipe',
'tipis', 'tips', 'to', 'toilet', 'total', 'tradisional', 'transit', 'transportasi',
'travel', 'triple', 'trus', 'tsb', 'tua', 'tub', 'tube', 'tudung', 'tugas', 'tuk',
'tukang', 'tulus', 'tunggu', 'tunjang', 'tunjuk', 'turis', 'turun', 'turut', 'tutup',
'tv', 'tvi', 'twin', 'type', 'uang', 'ubah', 'ubin', 'uda', 'udara', 'ukur', 'ukuran
ya', 'ulang', 'ulas', 'ultima', 'undang', 'undi', 'uneficient', 'uni', 'unik', 'univ
ersitas', 'unjung', 'untung', 'up', 'usang', 'utama', 'utara', 'utuh', 'variasi', 'v
ariatif', 'vegas', 'vegetarian', 'vendor', 'venue', 'very', 'via', 'vialita', 'vie
w', 'viewnya', 'vip', 'wajah', 'wajib', 'wangi', 'wanita', 'warga', 'waris', 'warn
a', 'waspada', 'wastafel', 'wedding', 'weird', 'wekday', 'welcome', 'western', 'wib',
'wifi', 'wilayah', 'wisata', 'wisatawan', 'with', 'wonton', 'wuenak', 'yah']
[[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]
...
[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]]

Menggunakan 5 Topik Teratas

In [7]:

```
# panggil class LDA
from sklearn.decomposition import LatentDirichletAllocation as LDA
```

```
n_topics = 5 #untuk mendapatkan jumlah topik terbaik perlu trial
lda = LDA(n_components=n_topics, learning_method='batch', random_state=0).fit(tf)
lda
```

Out[7]: LatentDirichletAllocation(n_components=5, random_state=0)

In [8]:

```
#training LDA
vsm_topics = lda.transform(tf)
#tampilkan hasil
print(vsm_topics.shape)
vsm_topics
```

(119, 5)

Out[8]:

```
array([[0.0068426 , 0.97287546, 0.00673103, 0.0067749 , 0.00677601],
       [0.00694955, 0.00693355, 0.00691056, 0.97228508, 0.00692126],
       [0.01273175, 0.01272259, 0.01265829, 0.94925665, 0.01263072],
       [0.00753287, 0.96989336, 0.00749834, 0.0074984 , 0.00757702],
       [0.00412958, 0.98350508, 0.004117 , 0.00412136, 0.00412697],
       [0.00331792, 0.00331809, 0.98672548, 0.00332555, 0.00331297],
       [0.00341551, 0.00343872, 0.00337813, 0.00338609, 0.98638155],
       [0.98576786, 0.00356885, 0.00354764, 0.00356062, 0.00355504],
       [0.01067223, 0.95729716, 0.01058828, 0.01079593, 0.0106464 ],
       [0.00731057, 0.0072451 , 0.00728716, 0.00726948, 0.97088768],
       [0.00614305, 0.97548814, 0.00608971, 0.00613432, 0.00614477],
       [0.00554387, 0.00550459, 0.00547327, 0.00553811, 0.97794016],
       [0.00425899, 0.00430235, 0.00424639, 0.98296963, 0.00422265],
       [0.00814929, 0.96748462, 0.00806166, 0.00814917, 0.00815525],
       [0.96289269, 0.00940847, 0.00921809, 0.00926466, 0.00921609],
       [0.00408965, 0.98374809, 0.00403119, 0.00405978, 0.0040713 ],
       [0.00422837, 0.00421763, 0.00420949, 0.98313075, 0.00421375],
       [0.96864196, 0.00788999, 0.00774518, 0.0078606 , 0.00786226],
       [0.00972057, 0.00982129, 0.00962898, 0.96118746, 0.00964169],
       [0.01126346, 0.01148561, 0.01135396, 0.95470861, 0.01118836],
       [0.00756947, 0.00758896, 0.96971571, 0.00755931, 0.00756655],
       [0.96755032, 0.00817443, 0.0080914 , 0.00808372, 0.00810012],
       [0.98516533, 0.00375124, 0.00366658, 0.00370056, 0.00371629],
       [0.00445875, 0.00443544, 0.00439471, 0.00445499, 0.98225611],
       [0.00657389, 0.00661985, 0.00652781, 0.9737209 , 0.00655755],
       [0.97085774, 0.00734805, 0.00724454, 0.00726926, 0.0072804 ],
       [0.00384252, 0.98469913, 0.00380494, 0.0038333 , 0.00382012],
       [0.00848716, 0.00853415, 0.00843784, 0.00846499, 0.96607588],
       [0.00399343, 0.98407833, 0.00396416, 0.00396704, 0.00399705],
       [0.01141258, 0.01145425, 0.01131927, 0.01142519, 0.95438871],
       [0.00377165, 0.00383616, 0.00373494, 0.9848663 , 0.00379096],
       [0.00682016, 0.97288092, 0.0067476 , 0.00679669, 0.00675463],
       [0.00403437, 0.62304051, 0.36495613, 0.00396911, 0.00399989],
       [0.9574962 , 0.01076333, 0.01058663, 0.0105892 , 0.01056465],
       [0.00601564, 0.00597088, 0.00600141, 0.97602491, 0.00598716],
       [0.00349433, 0.98604717, 0.00346135, 0.00349589, 0.00350125],
       [0.01067655, 0.95728987, 0.0106529 , 0.01067665, 0.01070403],
       [0.00623457, 0.00621834, 0.97514092, 0.00618712, 0.00621904],
       [0.00654983, 0.00657536, 0.00653488, 0.97379161, 0.00654832],
       [0.01351449, 0.94610124, 0.01337564, 0.01344637, 0.01356226],
       [0.9643879 , 0.00899861, 0.00881291, 0.00889893, 0.00890165],
       [0.98621162, 0.00347423, 0.00342249, 0.00343461, 0.00345706],
       [0.00965273, 0.96147933, 0.00959764, 0.00957233, 0.00969797],
       [0.98828435, 0.00292324, 0.00292682, 0.00293213, 0.00293346],
       [0.00484287, 0.9806466 , 0.00480611, 0.00486685, 0.00483757],
       [0.97457414, 0.0064287 , 0.00634715, 0.00632546, 0.00632455],
       [0.98804783, 0.00300977, 0.0029658 , 0.00299149, 0.00298512],
       [0.00971857, 0.51492552, 0.45602389, 0.0096116 , 0.00972042],
       [0.9880302 , 0.0029935 , 0.00298613, 0.00300057, 0.00298959],
       [0.9647782 , 0.00887789, 0.00875564, 0.00880197, 0.0087863 ]]
```


[0.98163393, 0.00459834, 0.0045941 , 0.00459828, 0.00457535],
[0.01363475, 0.84339516, 0.11551012, 0.01377765, 0.01368233],
[0.0045424 , 0.98190015, 0.00452681, 0.00452205, 0.0045086],
[0.00426173, 0.00424407, 0.00421627, 0.98302529, 0.00425265],
[0.00631721, 0.9746953 , 0.00633509, 0.00632329, 0.00632911],
[0.00855148, 0.00863818, 0.00840278, 0.00844803, 0.96595953],
[0.00665144, 0.97359348, 0.00658513, 0.00663477, 0.00653518],
[0.01018904, 0.0101998 , 0.01015258, 0.95929853, 0.01016004],
[0.00929003, 0.00938673, 0.96285765, 0.00923273, 0.00923286],
[0.00964037, 0.87048656, 0.00961509, 0.00967905, 0.10057893],
[0.00967563, 0.9615573 , 0.00956851, 0.00960676, 0.0095918],
[0.00811728, 0.9674475 , 0.00814979, 0.00818366, 0.00810177],
[0.00302907, 0.98789625, 0.00302117, 0.00302155, 0.00303197],
[0.97298922, 0.00677784, 0.0067241 , 0.00680515, 0.0067037],
[0.96966747, 0.00762676, 0.00754535, 0.00761522, 0.00754519],
[0.01084539, 0.95716884, 0.0106895 , 0.01062185, 0.01067442],
[0.00755364, 0.00762691, 0.00747297, 0.96978543, 0.00756105],
[0.00328536, 0.0032861 , 0.00327241, 0.98688461, 0.00327152],
[0.00320132, 0.00320651, 0.00315335, 0.00317898, 0.98725984],
[0.00440175, 0.9823256 , 0.00442903, 0.00443147, 0.00441215],
[0.0043825 , 0.00438686, 0.98246104, 0.00439236, 0.00437723],
[0.97675757, 0.00585813, 0.00578002, 0.00580203, 0.00580224],
[0.98694238, 0.00326297, 0.00326652, 0.00325697, 0.00327117],
[0.00349747, 0.98609138, 0.00345946, 0.00347865, 0.00347303],
[0.00338107, 0.00338667, 0.00336118, 0.98650976, 0.00336132],
[0.00812844, 0.96741633, 0.00814131, 0.00819532, 0.0081186],
[0.98439677, 0.0039478 , 0.00387067, 0.00389979, 0.00388497],
[0.00369296, 0.00369326, 0.00368334, 0.00369911, 0.98523133],
[0.02033622, 0.02024027, 0.02008688, 0.91906193, 0.0202747],
[0.00924491, 0.00942066, 0.00923674, 0.00925664, 0.96284105],
[0.0075223 , 0.96993414, 0.00747198, 0.00756534, 0.00750624],
[0.97208538, 0.0069799 , 0.00694441, 0.00701696, 0.00697335],
[0.00815605, 0.00821118, 0.96733571, 0.00817643, 0.00812063],
[0.00676468, 0.00677743, 0.00672546, 0.97298558, 0.00674685],
[0.96426214, 0.00891852, 0.00893198, 0.00902987, 0.00885749],
[0.00376241, 0.98496666, 0.00374201, 0.00376374, 0.00376518],
[0.00481296, 0.00481236, 0.00478912, 0.00480589, 0.98077966],
[0.00405492, 0.98383251, 0.00403445, 0.00404526, 0.00403286],
[0.00507778, 0.00513916, 0.0050556 , 0.97965893, 0.00506853],
[0.97462541, 0.00632946, 0.00633666, 0.00635899, 0.00634949],
[0.00351487, 0.00351245, 0.98597237, 0.00350389, 0.00349642],
[0.01838262, 0.01836532, 0.92649003, 0.01840865, 0.01835338],
[0.01406828, 0.94555182, 0.01345223, 0.01349232, 0.01343536],
[0.95460226, 0.01144902, 0.01130473, 0.01139123, 0.01125277],
[0.48042408, 0.48717351, 0.01077102, 0.01081336, 0.01081803],
[0.0080962 , 0.00806714, 0.00803998, 0.00807336, 0.96772333],
[0.005051 , 0.0050709 , 0.00504653, 0.00505945, 0.97977212],
[0.00750588, 0.00756288, 0.00746308, 0.00747296, 0.9699952],
[0.00474676, 0.00475854, 0.00473445, 0.00470002, 0.98106023],
[0.94603094, 0.01360695, 0.01341455, 0.01339164, 0.01355592],
[0.01358535, 0.01357784, 0.01359837, 0.94560292, 0.01363552],
[0.00537427, 0.00538219, 0.97848225, 0.00536128, 0.00540001],
[0.00366278, 0.98544935, 0.00363407, 0.00363289, 0.00362091],
[0.96573862, 0.00860775, 0.00856536, 0.00854171, 0.00854656],
[0.01204442, 0.20706148, 0.01208787, 0.01205702, 0.75674921],
[0.0031986 , 0.45768627, 0.53273212, 0.00319188, 0.00319113],
[0.00327443, 0.98689311, 0.0032669 , 0.00327786, 0.00328771],
[0.00536849, 0.00535982, 0.00535359, 0.97858014, 0.00533796],
[0.00631714, 0.00635604, 0.00630319, 0.00632829, 0.97469534],
[0.3565399 , 0.61547038, 0.0092349 , 0.00944594, 0.00930889],
[0.00359737, 0.98568945, 0.00358134, 0.00355549, 0.00357634],
[0.98196175, 0.00451439, 0.00454751, 0.0044948 , 0.00448154],
[0.00922876, 0.00924147, 0.00928656, 0.00928924, 0.96295397],
[0.00509442, 0.37261189, 0.00507066, 0.00505878, 0.61216426],

```
[0.00655067, 0.00654435, 0.97382389, 0.00651948, 0.00656162],
[0.00350452, 0.98596476, 0.00349305, 0.00353647, 0.0035012 ],
[0.00470644, 0.98116137, 0.00471399, 0.00470323, 0.00471496],
[0.01128298, 0.01144768, 0.01118628, 0.95484724, 0.01123581],
[0.98199277, 0.00452107, 0.00448768, 0.00450375, 0.00449472]])
```

In [9]:

```
#Tampilkan nilai-nilai setiap fitur
print(lda.components_)
```

```
[[1.19973002 0.20000145 0.20000352 ... 0.20000062 0.20000059 0.20000047]
 [3.19346328 4.19951679 2.19967556 ... 0.20000049 0.20000045 0.20000036]
 [0.20646458 0.20000278 0.20000683 ... 0.20000119 0.20000107 0.2000009 ]
 [2.20033899 0.20000204 1.19979728 ... 0.20000082 0.20000082 0.20000066]
 [0.20000312 0.20047695 0.20051681 ... 1.19999689 1.19999707 1.19999761]]
```

In [10]:

```
#hasil label topic model untuk setiap dokumen
import numpy as np
topics = np.argmax(vsm_topics, axis=1)
topics
```

Out[10]:

```
array([1, 3, 3, 1, 1, 2, 4, 0, 1, 4, 1, 4, 3, 1, 0, 1, 3, 0, 3, 3, 2, 0,
       0, 4, 3, 0, 1, 4, 1, 4, 3, 1, 1, 0, 3, 1, 1, 2, 3, 1, 0, 0, 1, 0,
       1, 0, 0, 1, 0, 0, 0, 1, 1, 3, 1, 4, 1, 3, 2, 1, 1, 1, 1, 0, 0, 1,
       3, 3, 4, 1, 2, 0, 0, 1, 3, 1, 0, 4, 3, 4, 1, 0, 2, 3, 0, 1, 4, 1,
       3, 0, 2, 2, 1, 0, 1, 4, 4, 4, 4, 0, 3, 2, 1, 0, 4, 2, 1, 3, 4, 1,
       1, 0, 4, 4, 2, 1, 1, 3, 0], dtype=int64)
```

In [11]:

```
#mencetak word fitur dengan nilai tertinggi pada setiap topik
n_top_words = 4 # jumlah fitur tertinggi yang kita tentukan
topic_words = {}
for topic, comp in enumerate(lda.components_):
    word_idx = np.argsort(comp)[::-1][:n_top_words]
    # store the words most relevant to the topic
    topic_words[topic] = [tf_vectorizer.get_feature_names()[i]+' '+str(comp[i]) for i in word_idx]
```

In [12]:

```
for topic, words in topic_words.items():
    print('Topic: %d' % topic)
    print(' %s' % ', '.join(words))
```

Topic: 0

kamar 44.95037351018805, inap 24.3022291144703, luas 22.48355057424015, bersih 15.23926144799837

Topic: 1

kamar 74.02180356596469, luas 41.376250696779515, kolam 26.53741444660403, renang 25.23551574225794

Topic: 2

kamar 12.779895885116645, kota 8.517582174429503, nyaman 7.496012733438991, renang 6.93207648352401

Topic: 3

kamar 28.20166167295463, renang 15.20331222682952, kolam 14.203328948262397, kota 12.2064099217546

Topic: 4

kamar 26.046265365774786, makan 13.835483418169021, renang 13.460268458267048, inap 12.539357819740447

Menggunakan 10 Topik Teratas

In [13]:

```
# panggil class LDA
from sklearn.decomposition import LatentDirichletAllocation as LDA
```

```
n_topics = 10 #untuk mendapatkan jumlah topik terbaik perlu trial
lda = LDA(n_components=n_topics, learning_method='batch', random_state=0).fit(tf)
lda
```

Out[13]: LatentDirichletAllocation(random_state=0)

In [14]:

```
#training LDA
vsm_topics = lda.transform(tf)
#tampilkan hasil
print(vsm_topics.shape)
vsm_topics
```

Out[14]:

```
(119, 10)
array([[0.00333435, 0.62140227, 0.00333382, ..., 0.00333408, 0.35192639,
        0.00333393],
       [0.00344856, 0.00344865, 0.00344833, ..., 0.00344863, 0.96896273,
        0.00344917],
       [0.00625093, 0.00625127, 0.00625151, ..., 0.00625097, 0.00625082,
        0.00625106],
       ...,
       [0.00232592, 0.00232594, 0.00232584, ..., 0.002326 , 0.00232586,
        0.00232602],
       [0.00555638, 0.00555666, 0.00555624, ..., 0.94999427, 0.00555615,
        0.00555622],
       [0.97999823, 0.00222251, 0.00222239, ..., 0.00222242, 0.00222245,
        0.00222238]])
```

In [15]:

```
#Tampilkan nilai-nilai setiap fitur
print(lda.components_)
```

```
[[1.1000014 0.1      0.1      ... 0.1      0.1      0.1      ]
 [1.09999787 1.09999584 1.09999923 ... 0.1      0.1      0.1      ]
 [0.1      0.1      0.1      ... 0.1      0.1      0.1      ]
 ...
 [1.09999835 0.1      0.1      ... 0.1      0.1      1.1      ]
 [1.10000218 0.1      0.1      ... 0.1      0.1      0.1      ]
 [1.09999891 0.1      0.1      ... 0.1      1.1      0.1      ]]
```

In [16]:

```
#hasil label topic model untuk setiap dokumen
import numpy as np
topics = np.argmax(vsm_topics, axis=1)
topics
```

Out[16]:

```
array([1, 8, 3, 1, 5, 5, 9, 7, 1, 9, 1, 4, 8, 1, 7, 1, 6, 9, 2, 5, 2, 0,
       7, 9, 7, 9, 1, 6, 1, 4, 5, 1, 5, 5, 4, 7, 1, 7, 9, 1, 9, 0, 1, 0,
       1, 7, 0, 6, 9, 8, 5, 7, 9, 7, 1, 4, 1, 3, 7, 6, 2, 9, 7, 0, 8, 7,
       7, 3, 4, 1, 8, 6, 0, 1, 3, 9, 8, 7, 3, 4, 1, 5, 2, 9, 7, 8, 4, 5,
       8, 0, 2, 3, 1, 8, 0, 6, 4, 4, 4, 7, 4, 2, 1, 0, 4, 5, 1, 5, 4, 7,
       7, 7, 4, 0, 7, 1, 6, 7, 0], dtype=int64)
```

In [17]:

```
#mencetak word fitur dengan nilai tertinggi pada setiap topik
n_top_words = 4 # jumlah fitur tertinggi yang kita tentukan
topic_words = {}
for topic, comp in enumerate(lda.components_):
    word_idx = np.argsort(comp)[::-1][:n_top_words]
    # store the words most relevant to the topic
    topic_words[topic] = [tf_vectorizer.get_feature_names()[i]+' '+str(comp[i]) for i in
```

```
In [18]: for topic, words in topic_words.items():  
        print('Topic: %d' % topic)  
        print(' %s' % ', '.join(words))
```

Topic: 0

kamar 18.870051966874158, inap 15.264265272749322, pilih 8.903652999334229, sarap 7.6194038396320645

Topic: 1

kamar 43.503063798994596, luas 19.80482972091912, bagus 18.677131551143503, makan 16.331279611050906

Topic: 2

renang 6.0555753767601095, kamar 5.099943820253244, kolam 5.042418857135264, nyaman 4.100035888895036

Topic: 3

kamar 9.995040009769426, renovasi 4.100041130896445, bersih 4.099988600275811, kolam 4.084015171057234

Topic: 4

kamar 13.132958391667241, makan 11.6745751348799, renang 11.670878162576814, kolam 9.638494903579396

Topic: 5

kamar 14.099991030069608, nyaman 10.100050836282408, kota 9.10006903865884, gunung 7.078459306767805

Topic: 6

kamar 8.191300262476082, layan 6.100067339597893, bersih 5.100049392851896, ramah 4.100039833155047

Topic: 7

kamar 30.468491662529395, luas 17.812403878173903, kolam 12.09997659338341, renang 12.099968505872127

Topic: 8

kamar 18.853741083217606, luas 7.682815097030212, bersih 6.79808943147419, bangun 6.752728708655957

Topic: 9

kamar 23.785417974147364, renang 11.857430731767646, luas 10.887490368555644, kolam 10.87187042696922