

Vývoj a pilotní odzkoušení webové aplikace pro návrh scénářů budoucího vývoje

Vícerozměrné a grafické analýzy výstupů empirických výzkumů metodami strojového učení.

Příloha 2 k dílčímu výstupu klíčové aktivity

Aktualizace strategického plánu se zapojením veřejnosti

projektu

PŘÍVĚTIVÝ INOVATIVNÍ ÚŘAD STRMILOV

(CZ.03.4.74/0.0/0.0/18_092/0014301)



spolufinancovaného Evropskou unií prostřednictvím Ministerstva práce a sociálních věcí.

Řešitelé:

doc. Dr. Ing. Jan Voráček, CSc., Katedra technických studií

Mgr. Martina Černá, PhD., Katedra sociální práce

Vysoká škola polytechnická Jihlava, leden 2022



MINISTERSTVO PRÁCE
A SOCIÁLNÍCH VĚCÍ



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Úvod

Tento dokument shrnuje výsledky vícerozměrných analýz následujících empirických výzkumů veřejného mínění:

- a) **Život v obcích Strmilovska:** elektronická (<https://bit.ly/2QJXH2D>) i fyzická anketa, realizovaná od května do srpna 2020. Úplné znění otázek i jejich základní vyhodnocení je na webu projektu PIUS: <https://bit.ly/3vtkiGd>.
- b) **Rozhovory s mladými lidmi Strmilovska:** individuální rozhovory s ochotnými a momentálně dostupnými respondenty, odpovídajícími požadovanému výzkumnému profilu, realizovaná od května do srpna 2020. Úplné znění otázek i jejich základní vyhodnocení je jak na webu projektu PIUS (<https://bit.ly/3m0BG0t>), tak také v dokumentu na adrese: <https://bit.ly/3rBWGMw>.
- c) **Hledání zájmu a formy podpory starostů/starostek malých obcí při rozvoji jejich komunity:** elektronická anketa (<https://bit.ly/3sszdP5>), realizovaná v březnu 2021.

Metodika

Tam, kde to bylo možné, byly při zpracování odpovědí použity metody strojového učení. Jejich výhodou je, že se neomezují pouze na vizualizaci či kvantifikaci výsledků jediné otázky nebo vzájemného vztahu dvou proměnných, ale jsou schopny identifikovat i obecnější souvislosti, platné pro celý soubor položených dotazů. Na druhé straně ale bývá obvyklé, že podaří-li se takovéto vazby najít, nebývá jich mnoho a často odpovídají běžnému chápání analyzované problematiky. Přidanou hodnotou metod strojového učení je pak alespoň explicitní validace empirických hypotéz.

Analyzované ankety obsahovaly také několik tzv. otevřených odpovědí, jimiž mohli respondenti vyjádřit libovolné vlastní názory. Tyto nestrukturované informace lze buď dodatečně kategorizovat a analyzovat společně se zbytkem souboru, nebo interaktivně zpracovat některým z textových analyzátorů. V této zprávě je ukázána pouze přímočará vizualizace neočištěných textů formou slovních grafů (word cloud), kdy velikost písma odpovídá počtu výskytů daného slova. Přestože jde jen o velmi přibližnou aproximaci, určité závěry lze vyvodit i z ní.

Příprava dat

Strojové učení bylo aplikováno pouze na vybrané výstupy anket s občany Strmilovska a se starosty/starostkami menších sídel. Údaje, získané z cílených rozhovorů s mladými lidmi byly pouze orientačně vizualizovány, protože jde o relativně specifický populační vzorek malého rozsahu, zkoumaný odlišnou metodikou. Znění i kódování otázek pro obě vícerozměrné analýzy je v Přílohách 1 a 2 tohoto dokumentu.

1.1.1 Předzpracování ankety s občany Strmilovska

- Otevřené otázky číslo 16 až 18, týkající se kulturního domu, byly jako příliš specifické z analýz vyřazeny zcela.
- Z dlouhých otázek s vícenásobnou možností odpovědi a možnou vlastní variantou (otázky číslo 2, 11, 12, 19 a 21) byly vybrány nejčastější odpovědi, typicky tři nebo čtyři, a těm byly přiřazeny nové kategorie, značené písmeny (např. 2a, 12d apod.). Tím vzrostl počet analyzovaných otázek v každém záznamu z 23 na 43.
- Chybějící odpovědi v předdefinovaných variantách otázek byly nahrazeny nulami. Nulami byly doplněny i nevyužité možnosti u vícenásobných otázek, takže nulovou hodnotu bylo třeba interpretovat vždy v kontextu otázky. Ve druhém případě totiž nejde o vůli nebo preference respondenta, ale o jeho konkrétní vymezení vůči většinové části skupiny.
- Z takto upravených dat byly pořízeny následující dva soubory:
 - a) **Redukovaná data**, ze kterých byly odstraněny všechny neúplné záznamy. Tím se analyzovaný soubor zmenšil z původních 55 na 23 záznamů a nuly se v něm mohou vyskytnout pouze u zkrácených vícenásobných otázek s písmenovou příponou.
 - b) **Kompletní data**, čítající 55 záznamů, z nichž řada je ale různým způsobem neúplná, což bylo opět reprezentováno nulami.

Je zřejmé, že oba analyzované soubory byly zatíženy systematickou chybou (bias), ale výzkumný předpoklad byl, že nejvýznamnější závěry by se měly komplementárně promítnout do každého z nich a budou proto identifikovány při závěrečném expertním zpracování dílčích výstupů. Za základní zdroj informací byla považována čistá redukovaná data a zjištěné významné relace byly dále křížově validovány prostřednictvím kompletního souboru.

1.1.2 Předzpracování ankety se starosty

- Ze získaných dat byly odebrány všechny otevřené otázky a neúplné záznamy. Výsledný soubor pak obsahoval 58 záznamů, z nichž každý měl 7 kategoriálních otázek. Na kategorie byly formou ekvalizace histogramu převedeny i číselné odpovědi typu věk nebo počet obyvatel, protože se tím zjednoduší struktura generovaných analyzátorů.

Provedené analýzy

Obě ankety byly zpracovány prostřednictvím následujících dvou typů úloh:

1.1.3 Redukce počtu zpracovávaných proměnných

Cílem této techniky bylo nalezení takové nejmenší sady proměnných (prediktorů), která extremalizuje zvolenou vzdálenostní metriku vzhledem k vybrané výstupní třídě „*Spokojený život na Strmilovsku*“ a „*Jak hodnotíte sousedské vztahy v obci?*“. Jde tedy o úlohu příbuznou klasifikaci neboli učení s učitelem, ale výsledkem zde není návrh účinného vstup-výstupního mapování, ale pouze eliminace méně významných atributů. Tímto způsobem zjistíme, které z otázek nejvíce souvisí se spokojeností obyvatel, ať už v kladném nebo v záporném smyslu. Výpočty byly realizovány v prostředí analytického nástroje *SAS Enterprise Miner* vhodně parametrizovanými funkcemi *StatExplore* a *HP Variable Selection* v implicitním nastavení, které na základě odlišných principů hledají nejmenší sadu proměnných, maximalizující celkový rozptyl dat vzhledem k výstupu.

V případech, kdy citlivost těchto metod nestačila, byly pro detailnější nalezení rozdílů využity kvalitativně odlišné metody *Gradient Boosting* (GB) a *Random Forest* (RF), založené na principu sekvenční (GB) a paralelní (RF) konstrukce sady rozhodovacích stromů. Tím se úloha paralelizuje a finální výsledky integrují širší spektrum dílčích výstupů.

1.1.4 Shluková analýza na úrovni proměnných

Jednalo se o tzv. učení bez učitele, implementované vhodně nastavenou funkcí *Variable Clustering*. Ta k sobě bez jakékoli další apriorní informace přiřazuje proměnné, blízké ve smyslu zvoleného kritéria. Tím byla i v tomto případě maximalizace celkového rozptylu zdrojových dat minimálním počtem vytvořených shluků. Shlukováním proměnných lze zjistit i relace dílčích hodnot jednotlivých kategorií, a to jak mezi sebou, tak i vzhledem ke zbývajícím proměnným. Výsledky, získané shlukováním, je ale třeba vždy expertně posoudit a diskutovat.

Výsledky

Život v obcích Strmilovska

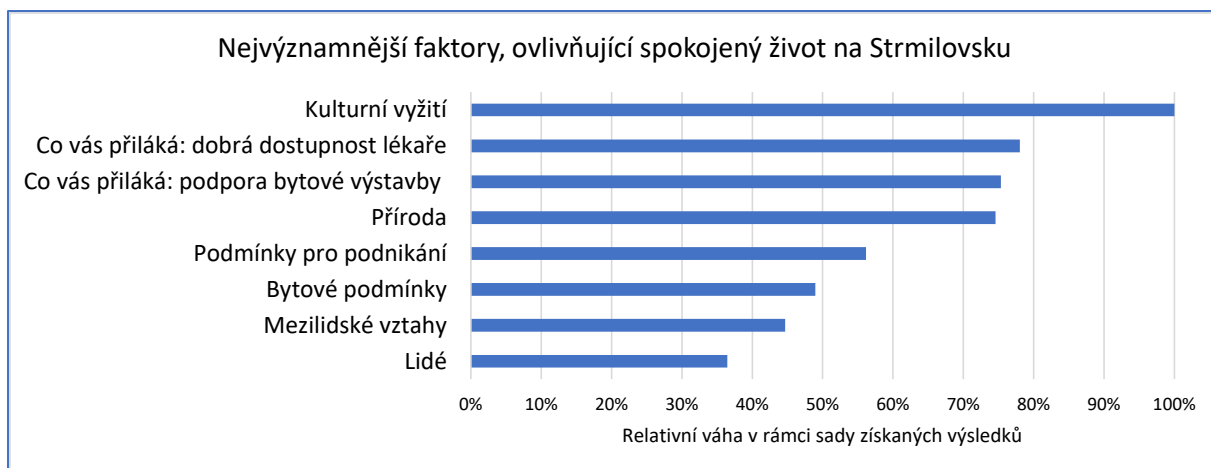
1.1.5 Vícerozměrné analýzy

Jednotlivé odpovědi z tohoto průzkumu byly posuzovány z hlediska míry jejich celkové vazby k první otázce „*Žijete ve své obci spokojený život?*“. Vytváření podmínek pro spokojený život občanů je totiž jedním z obecných cílů městského úřadu a přívětivost, reprezentovaná ovzduším vzájemné důvěry a spolupráce představuje jednu z účinných strategií jeho naplňování. Díky participativnímu zapojení spokojených občanů lze totiž předpokládat, že dlouhodobé rozvojové iniciativy obce budou optimálním vyjádřením společného zájmu a nikoli jen reflexivním naplňováním nerevidovaných historických záměrů, administrativní implementací legislativních či politických direktiv nebo výslednicí momentálních tlaků vnějšího prostředí.

Provedená vícerozměrná analýza hledala odpovědi na následující otázky:

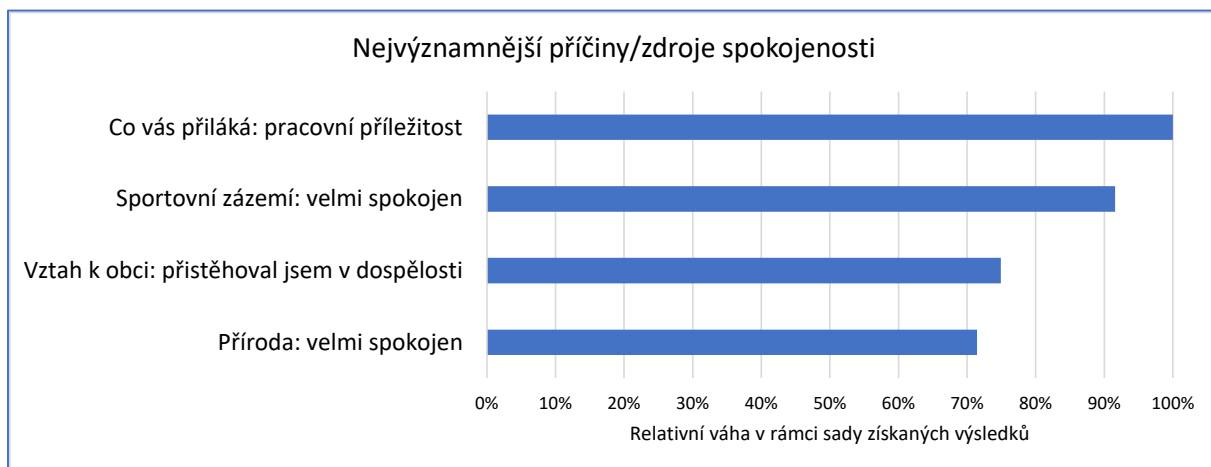
- a) Jaké jsou nejvýznamnější faktory, ovlivňující spokojený život Strmilovska?**
- b) Jaké jsou nejvýznamnější příčiny nebo zdroje spokojenosti občanů Strmilovska?**
- c) Jaké jsou nejvýznamnější příčiny nebo zdroje nespokojenosti občanů Strmilovska?**

Získané výsledky jsou ukázány v grafech na obrázcích 1 – 3 a také jsou pod nimi vzorově interpretovány. Přitom je třeba mít na paměti, že nalezené faktory jsou vzájemně nezávislé, tj. není nezbytné, aby platily zároveň. Pro vyvozované závěry je důležité i to, že analyzovaný soubor je malý a velmi pravděpodobně není ani reprezentativní, tedy rovnoměrně zahrnující všechny typické skupiny místních občanů. Proto doporučujeme vnímat prezentované závěry spíše z hlediska metodického a v případě zájmu o seriózní datové výstupy současný soubor odpovědí dále rozšiřovat.



Obrázek 1 Nejvýznamnější faktory, ovlivňující spokojený život Strmilovska

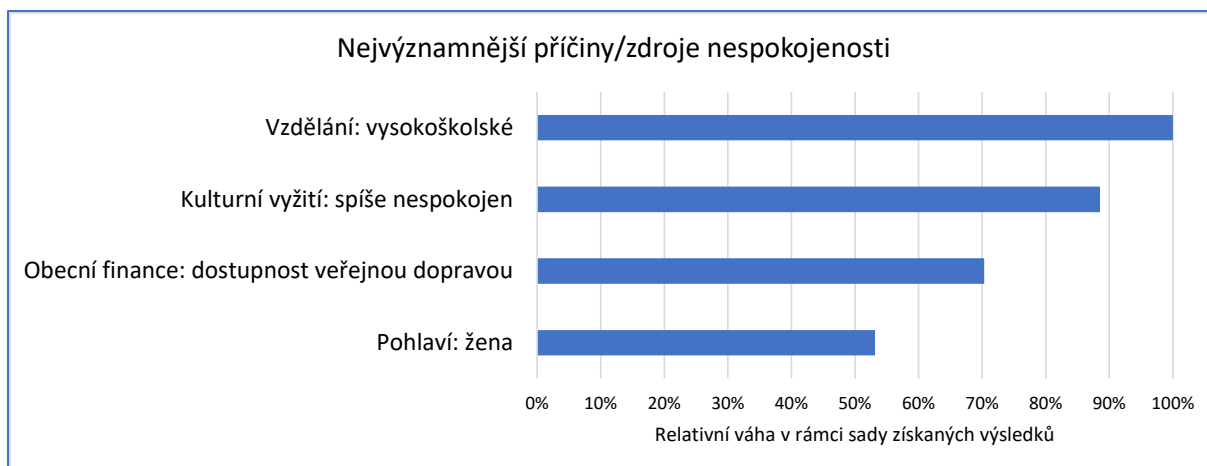
Výsledné faktory z obrázku 1 se vztahují souborně ke všem variantám odpovědí na zkoumanou otázku. Pokud tedy vzorový strmilovský respondent považuje za nejdůležitější součást svého spokojeného života kulturní vyžití znamená to, že lidé, vedoucí spokojený život jsou s místním kulturním vyžitím buď velmi spokojeni nebo spokojeni, zatímco méně či více nespokojeným občanům zásadně nevyhovuje ani úroveň kulturního vyžití. Podobně spokojený občan oceňuje nebo by ocenil dostupnost lékaře a podporu bytové výstavby, zatímco nespokojený považuje obě tyto oblasti za nedostatečně rozvinuté. Příroda je sice vnímána spokojenými i nespokojenými obyvateli pozitivně, avšak skupina spokojených ji hodnotí velmi vysoko, zatímco méně spokojení se přiklání spíše k průměru. Podobná rozložení preferencí je také u bydlení, mezilidských vztahů a sociálních vazeb, reprezentovaných faktorem Lidé. Podmínky pro podnikání jsou sice také rozprostřeny pouze do poloviny spektra nabízených možností, ale tentokrát v opačném smyslu – spokojení občané je považují za průměrné, zatímco nespokojení za méně vyhovující.



Obrázek 2 Nejvýznamnější příčiny/zdroje spokojenosti občanů Strmilovska.

Obrázek 2, ilustrující nejvýznamnější příčiny či zdroje spokojenosti občanů Strmilovska, se na rozdíl od předchozího kumulativního rozboru soustředí pouze na skupinu spokojených občanů. Ukazuje, že tito obyvatelé jednoznačně nejvíce reflektují na pracovní příležitosti v místě, jsou spokojeni s přírodou i sportovním vyžitím a obvykle se nejedná o místní rodáky.

Obrázek 3 říká, že nespokojení se ve Strmilově cítí oslovení vysokoškoláci, dále lidé, kterým zde chybí kvalitní kulturní zázemí a také ti, kteří vidí rezervy v místní dopravní obslužnosti. Zajímavá je i prokazatelně vyšší míra nespokojenosti žen, které se dotazníkového šetření zúčastnily.



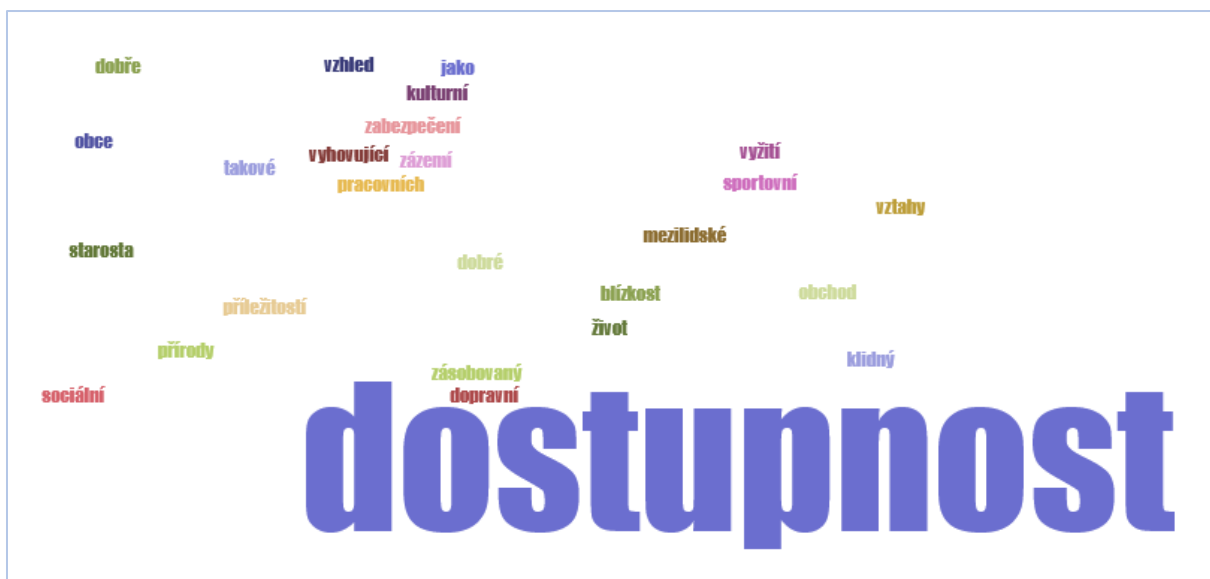
Obrázek 3 Nejvýznamnější příčiny/zdroje nespokojenosti občanů Strmilovska.

1.1.6 Grafické výstupy ankety s obyvateli Strmilovska

Tato forma zpracování je natolik názorná, že není třeba ji komentovat. Velikost textu reflektuje pouze počet jeho výskytů, nikoli jejich kontext – viz např. termín dostupnost, který se vyskytuje v opačně postavených otázkách na obrázcích 4 i 5. Z obrázku 5 je zřejmý i celkově nižší počet odpovědí.



Obrázek 4 Textový graf odpovědí na otázku *Co vás ve Strmilově trápí?*



Obrázek 5 Textový graf odpovědí na otázku *Co se vám v obci líbí?*

Rozhovory s mladými lidmi Strmilovska

K tomuto průzkumu byly jako pseudo-vícerozměrný doplněk k výše odkazované tradiční formě zpracování doplněny pouze textové grafy na obrázcích 6 – 13. Záměrně jsou ponechány bez komentáře, neboť ten by byl u tak malého vzorku a zvolené formy vyhodnocení spíše spekulativní.



Obrázek 6 Textový graf odpovědí na otázku *Co se vám v obci líbí?*



Obrázek 7 Textový graf odpovědí na otázku *Co se vám v obci nelíbí?*



Obrázek 8 Textový graf odpovědí na otázku *Co budete nejméně potřebovat v nejbližších pěti letech?*

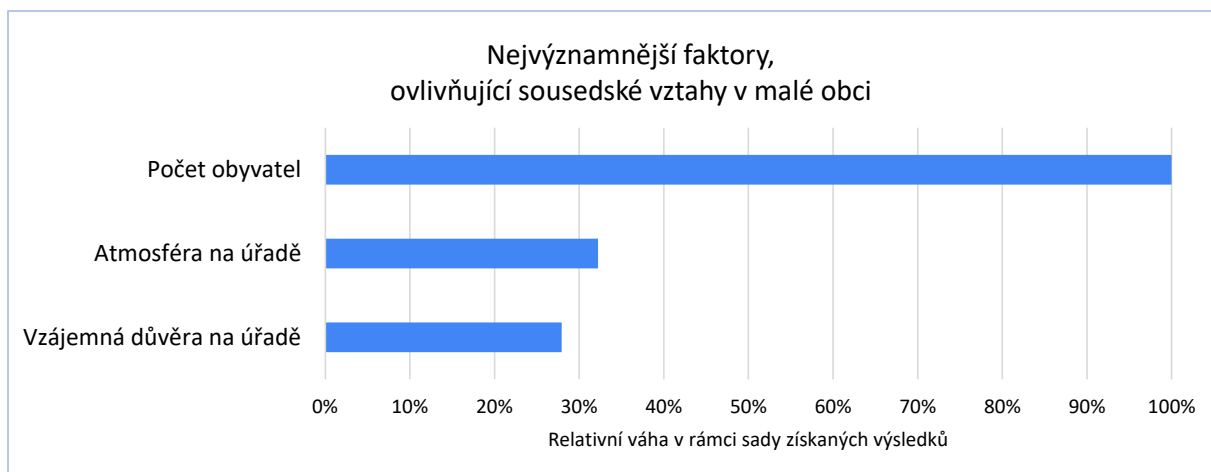




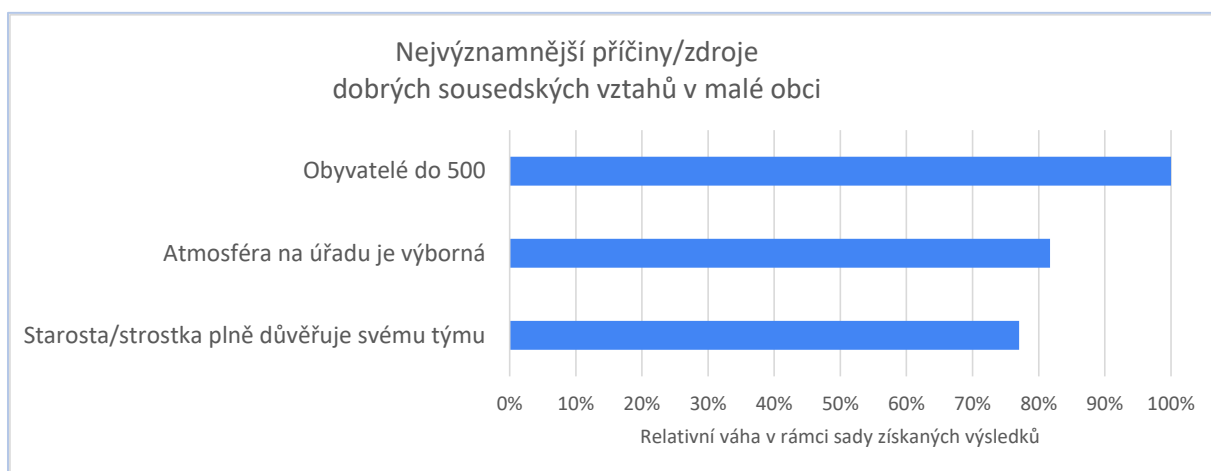
Obrázek 11 Textový graf odpovědí na otázku *Myslíte si, že obyvatelé obce mají dostatek příležitostí ke vzájemným kontaktům?*



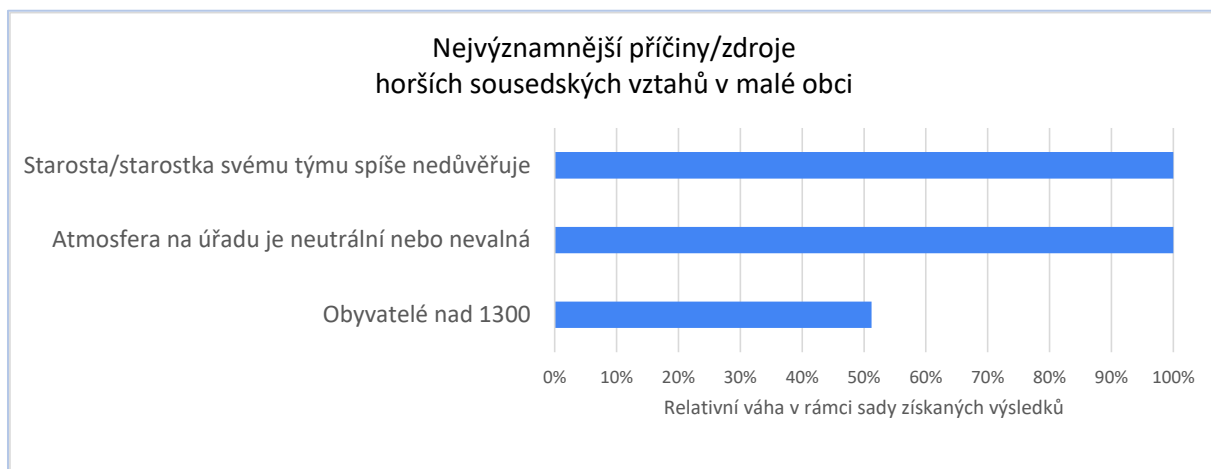
Obrázek 12 Textový graf odpovědí na otázku *Budoucí účel budovy kulturního domu?*



Obrázek 14 Nejvýznamnější faktory, ovlivňující sousedské vztahy v malé obci



Obrázek 15 Nejvýznamnější příčiny nebo zdroje dobrých sousedských vztahů v malé obci



Obrázek 16 Nejvýznamnější příčiny nebo zdroje horších sousedských vztahů v malé obci

[illegible]

Obrázek 18 Textový graf odpovědí na otázku *Oblasti a formy participace?*

Příloha 1: Otázky z průzkumu Život v obcích Strmilovska**Tabulka 1** Sada anketních otázek pro občany, zpracovaná postupy strojového učení

ID	Znění	Kód	Možnosti
1	Žijete ve své obci spokojený život?	1	Jednoznačně ano
		2	Spíše ano
		3	Spíše ne
		4	Jednoznačně ne
		5	Nedovedu posoudit
2	Co se vám v obci líbí?	a	Blízkost přírody
		b	Klidný život
		c	Vzhled obce jako takové
		d	Dobré mezilidské vztahy
3	Lidé (rodina, sousedé, kamarádi a známí)	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen
4	Mezilidské vztahy	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen
5	Kulturní vyžití	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen
6	Pracovní příležitosti	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen
7	Sportovní zázemí	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen

8	Příroda	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen
9	Bytové podmínky	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen
10	Podmínky pro podnikání	1	Velmi nespokojen
		2	Spíše nespokojen
		3	Ani spokojen ani nespokojen
		4	Spíše spokojen
		5	Velmi spokojen
11	Co vás ve Strmilově trápí?	a	Nedostatek pracovních příležitostí
		b	Nevyhovující veřejná doprava
		c	Špatná dostupnost obchodů a služeb
		d	Špatné vztahy mezi lidmi

12	Kde se dozvídáte aktuální informace od obce?	a	Oficiální webové stránky
		b	Místní rozhlas
		c	Strmilovské noviny
		d	FB stránka "Strmilov - pěkné místo k životu"
13	Sledujete informace o dění v obci na oficiálních webových stránkách?	1	Pravidelně - 1x týdně
		2	Občas - 1x měsíčně
		3	Nesleduji
		4	Nemám internetové připojení
14	Sledujete informace o dění v obci na Facebook stránce obce "Strmilov-pěkné místo k životu"?	1	Pravidelně - 1x týdně
		2	Občas - 1x měsíčně
		3	Nesleduji
15	Myslíte si, že obyvatelé obce mají dostatek příležitostí ke vzájemným společenským kontaktům?	1	Jednoznačně ano
		2	Spíše ano
		3	Spíše ne
		4	Jednoznačně ne
		5	Nedovedu posoudit
19	Co vás přiláká do Strmilova, abyste se zde usadili a založili zde rodinu?	a	Dobrá dostupnost do škol
		b	Vhodná pracovní příležitost
		c	Dobrá dostupnost lékaře
		d	Možnost koupě pozemku
20	Pokud již v obci žijete, plánujete zůstat do konce života?	1	Jednoznačně ano
		2	Spíše ano
		3	Spíše ne
		4	Jednoznačně ne
		5	Nedovedu posoudit
21	Na co byste primárně využili finance obce?	a	Podpora lepší zdravotní a sociální péče
		b	Podpora bytové výstavby
		c	Zajištění lepší dostupnosti obce veřejnou dopravou
		d	Zřízení dalších provozoven obchodu a služeb v obci
22	Pohlaví	1	Muž
		2	Žena
23	Věk	1	15-21

		2	22-29
		3	30-49
		4	50-64
		5	65 a více
24	Vaše vzdělání	1	Základní
		2	Střední odborné
		3	Střední odborné s maturitou
		4	Vyšší odborné
		5	Vysokoškolské
25	Bydlím	1	Ve Strmilově
		2	V České Olešné
		3	V Palupíně
		4	V Malém Jeníkově
		5	V Leštině
		6	Nebydlím, mám zde chalupu
		7	V blízké vesnici
26	Vztah k obci	1	Žiji zde od narození
		2	Přistěhoval jsem se sem v dětství
		3	Přistěhoval jsem se sem až v dospělosti
		4	Mám zde chalupu
		5	Mám zde rodinu
		6	Chci zde zestárnout

Příloha 2: Otázky z průzkumu Hledání zájmu a formy podpory starostů/starostek malých obcí při rozvoji jejich komunity

Tabulka 2 Sada anketních otázek pro starosty, zpracovaná postupy strojového učení

ID	Znění	Kód	Možnosti
1	Jak hodnotíte sousedské vztahy v obci?	1	Výborné
		2	Spíše dobré
		3	Dobré
		4	Nevím
		5	Spíše špatné
		6	Špatné
		7	Hrozné
2	Co je podle Vás důležité, aby obec rozkvetla bohatým kulturně-společenským životem?	Otevřená otázka	
3	Jaká je atmosféra na úřadě, který vedete?	1	Výborná
		2	Dobrá
		3	Neutrální
		4	Nevalná
		5	Nepříjemná
4	Důvěřujete týmu svého úřadu?	1	Ano
		2	Spíše ano
		3	Nevím
		4	Spíše ne
		5	Ne
5	Dáváte prostor občanům k tzv. participaci? (tj. zapojování veřejnosti do rozhodování o věcech v obci v období mimo volby) Pokud ano, jaké? Pokud ne, proč?	Otevřená otázka	
6	Zajímá Vás osobní rozvoj?"	1	Ano
		2	Ne
7	Pokud Vás osobní rozvoj, který by posílil Vaší motivaci při rozvoji komunity, zajímá, v jaké oblasti chcete na sobě pracovat?	Otevřená otázka	
8	Pohlaví	1	Muž
		2	Žena

9	Věk	1	do 43
		2	43-48
		3	48-53
		4	53-58
		5	nad 58
10	Počet obyvatel obce, kterou vedete:	1	do 500
		2	500-900
		3	900-1100
		4	1100-1300
		5	1300-1600
		6	1600-2000